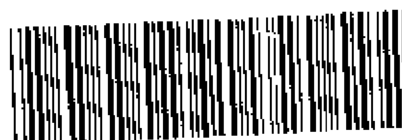
南开大学数学教学丛书

概 率 论

杨振明 著





科 学 出 版 社

1 9 9 9

内 容 简 介

本书为南开大学数学系教材.内容包括:事件与概率;随机变量;数字特征与特征函数;极限定理等.本书是作者多年教学工作经验的总结,内容丰富,深入浅出,论述严谨,每一节后都有习题,有助于读者理解书中内容.

本书可供高等学校数学系学生和教师阅读和参考.

图书在版编目(CIP)数据

概率论/杨振明著.-北京:科学出版社,1999.1
(南开大学数学教学丛书)
ISBN 7-03-006746-0

Ⅰ.概… Ⅱ.杨… Ⅲ.概率论 Ⅳ.O211

中国版本图书馆CIP数据核字(97)第12215号

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717

新蕾印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1999年1月第一版 开本:850×1168 1/32
1999年1月第一次印刷 印张:8 5/8
印数:1-2800 字数:231 000

定价:14.50 元

(如有印装质量问题,我社负责调换(环伟))

序

海内外华夏炎黄子孙都盼望中国早日成为数学大国，也就是“实现中国数学的平等和独立”[1]。平等和独立是由中国出类拔萃的数学家及其杰出的研究工作来体现的，要有出类拔萃的数学家就要培养一批优秀的研究生、大学生。这批人不在多，而在精，要层次高。也就是要求他们热爱数学、基础扎实、知识面广、能力强。

80 年代中期，国家采纳了陈省身先生的几个建议。建议之一是为培养高质量的数学专业的大学生，需要建立数学专业的试点班．经过胡国定先生等的努力，1986 年在南开大学建立了数学专业的试点班。这些作法取得了成功，并在基础学科的教学中有了推广。1990 年在全国建立“国家理科基础学科研究和教学人才培养基地”。其后南开大学数学专业成为基地之一。从 1986 年到现在的 10 余年中南开数学专业是有成绩的，例如他们四次参加全国大学生数学竞赛获三次团体第一，一次团体第三．在全国和国际大学生数学建模比赛中多次获一等奖。毕业生中的百分之八十继续攻读研究生，其中许多人取得了很好的成绩。

当然，取得这些成绩是与陈省身先生的指导、帮助分不开，是与国内外同行们的支持与帮助分不开的。如杨忠道，王叔平，许以超，虞言林，李克正等或参与教学计划、课程设置、课程内容的制订，或到南开任教等等。有了这些指导、帮助与支持，南开基础数学专业得以广泛吸收国内外先进的数学教学经验，并以此为基础对数学教学进行了许多改革、创新。

这套丛书是南开大学数学专业的部分教材，编著者们长期在南开数学专业任教，不断地把自己的心得体会揉和到基础知识和基本理论的讲述中去，日积月累地形成了这套教材。所以可以说这

[1] 陈省身：《在“二十一世纪中国数学展望”学术讨论会开幕式上的讲话》

些教材不是“编”出来的，而是在长期教学中“教”出来的，“改”出来的，凝聚了我们的一点心血。这些教材的共同点，也是我们教学所遵循的共同点是：首先要加强基础知识、基础理论和基本方法的教学；同时又要适当地开拓知识面，尤其注意反映学科前沿的成就、观点和方法；教学的目的是丰富学生的知识与提高学生的能力，因此配置的习题中多数是为了巩固知识和训练基本方法，也有一些习题是为训练学生解题技巧与钻研数学的能力。

我们要感谢中国科学出版社主动提出将这套教材出版。这对编著者是件大好事。编著者虽然尽了很大努力，但一则由于编著者的水平所限，二则数学的教育和所有学科的教育一样是在不断发展之中，因此这套教材中缺欠和不足肯定存在。我们诚挚希望各位同行不吝指正，从而使编著者更明确了解教材及教学中的短长，进而扬长避短，改进我们的教学。同时通过这套教材也可向同行们介绍南开的教学经验以供他们参考，或许有益于他们的工作。

我们再次感谢帮助过南开的前辈、同行们，同时也希望能继续得到他们和各位同行的帮助。办好南开的数学专业，办好所有学校的数学专业，把中国数学搞上去，使中国成为数学大国是我们的共同愿望！这个愿望一定能实现！

全体编著者

于南开大学

1998 年 6 月

目　　录

第一章　事件与概率

第二章　随机变量

第三章　数字特征与特征函数

第一章　事件与概率

1.1　基本概念

1.1.1　随机现象

客观世界中存在着两类现象，一类是在一定条件下必然出现的现象，称之为**必然现象**；另一类是在一定条件下可能出现也可能不出现的现象，称作**随机现象**.

在标准大气压下，100° C 的纯水必然沸腾；带异性电荷的小球必然相互吸引；在恒力作用下的物体必然作匀加速运动等等都是必然现象的例子. 以往的各种数学学科的主要内容就是研究必然现象中的数量规律. 例如，在重力作用下，物体的位移随时间变化的函数 $x(t)$，就由二阶微分方程 $x''(t)=g$ 来描述，其中 g 为重力加速度.

随机现象也是广泛存在的. 例如，抛掷一枚硬币，它可能是正面朝上也可能是背面朝上，就是说，“正面朝上”这个结果可能出现也可能不出现；记录一段时间内某电话机的呼唤次数可能较多也可能较少；下一个交易日股市的指数可能上升也可能下跌，而且升跌幅度的大小也不能事先确定. 这些都是随机现象的例子. 许多影响事物发展的偶然因素的存在，是产生随机现象中不确定性的原因. 例如，股市指数的变化取决于金融政策的变化，上市企业的经济状况，股民的经营行为以及其它国家的股市沉浮等诸多不确定因素. 这些因素发展变化的偶然性，决定了股市升跌的随机性.

虽然随机现象中出现什么样的结果不能完全预言，但是可以假定全部可能结果是已知的. 在上述例子中，抛一枚硬币只会有“正面”与“背面”这两个可能结果，电话呼唤次数必定是某个非负整数，而股指的升跌幅度充其量假定它可能是任意的实数. 可见“全部可能结果的集合是已知的”这个假定是合理的，而且它会给我们带来许多方

便.

进行一次试验，如果其所得结果不能完全预言，但其全体可能结果是已知的，则称此试验为**随机试验**．这里及今后所使用的“试验”这一术语，其含义是广泛的．它既可以是通常意义下的物理或化学的实验，也可以是对自然或社会现象的观测与记录.

表 1.1 抛硬币试验资料

试验者	抛硬币次数	出现正面次数	出现正面频率
蒲丰	4040	2048	0.5069
德莫根	4092	2048	0.5005
费勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005
罗曼诺夫斯基	80640	39699	0.4923

虽然一次随机试验的结果不能完全预言，但是，在相同条件下大量重复此试验时，则会呈现出一定的数量规律性．这一点被历史上许多人的试验结果所证明．表 1.1 列出了蒲丰等人连续抛掷均匀硬币所得的结果．从表中的数据可以看到，当抛掷次数很大时，正面出现的频率非常接近 0.5，就是说，出现正面与出现背面的机会差不多各占一半.

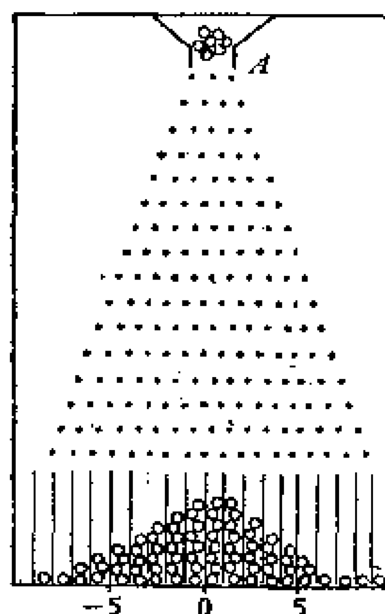


图 1.1 高尔顿板

另一个著名的例子是高尔顿钉板试验．在一块竖起的木板上钉上一排排间隔相等的铁钉 (见图 1.1)．每一排各个钉子正好对准上面一排两个相邻铁钉的中央．这样，当小球从两钉之间的间隙下落时，由于下一排铁钉的碰撞，它将以相等的可能性向左或向右落下．小球再走过两钉间隙时，又遇到下一排铁钉的碰撞．如此下去，当小球走过最后

一排铁钉的间隙后，便落入下方的被分割成条状的容器中．让许许多多小球自上而下地重复这一下落过程．尽管每次指定小球下落的位置无法预言，但许多小球堆积的边缘轮廓线总是两头矮中间高且左右近于对称的钟形曲线．就是说，落入各条状容器内小球数所占的比例基本保持不变．

上面列举的两个试验的结果表明，在相同条件下大量地重复某一随机试验时，各可能结果出现的频率稳定在某个确定的数值附近．称这种性质为**频率的稳定性**．频率稳定性的存在，标志着随机现象也有它的数量规律性．**概率论就是研究随机现象中数量规律的数学学科**．

历史上概率论起源于赌博．三百多年以前，现代化生产与技术尚处于萌芽状态，对自然界中随机现象规律性的研讨还没有提到议事日程．但此时，具有悠久历史的游戏与赌博却发达得多．人们借助于骰子，纸牌以及形形色色的工具进行赌博，遇到了许多无法解释的问题．由于输赢是无法预言的，而且涉及到金钱的得失，所以了解其中的数量规律就变得越来越迫切了． 17 世纪中叶，赌徒中一些有身分的人开始向他们的数学家朋友请教．当时在欧洲颇具声望的数学家费尔马、巴斯卡、惠更斯等人都参加了有关的讨论．于是产生了概率论的基本概念，并给出了计算“等可能型”概率的一套方法．

19 世纪末至 20 世纪初，在现代工业技术蓬勃发展的大潮中，概率论也取得了飞速的发展．特别是柯尔莫戈洛夫等人建立了概率论的公理化体系，奠定了概率论的严格的数学基础，也沟通了概率论与现代数学中其它分支之间的联系．近年来，概率论已被广泛应用于自然科学，工程技术，经济理论，经营管理等许多方面．特别是对金融领域中随机现象的研究与应用有了长足的发展，形成了“金融数学”的重要组成部分．概率论这个有特色的数学分支呈现出蓬勃发展的局面．当前，我国高等学校的许多专业都开设程度不同的概率论课程，单从这一点就可看到概率论作为一门基础学科在社会发展中所起的巨大作用．

最后我们指出，在今天的概率论教材中，仍广泛采用“抛硬币”，“掷骰子”等与赌博有关的例子．除了一些典型例题有很重要的历史

地位外，主要是由于这些例子所涉及的概念简单，规则明了，易于阐明概率论的基本概念与方法.

1.1.2 样本空间

随机试验的每个可能结果称为一个**样本点**，全体样本点所组成的集合称为**样本空间**. 习惯上分别用小写的 ω 与大写的 Ω 表示样本点与样本空间. 如前所述，我们总假设样本空间 Ω 是确定的.

例 1.1 抛两枚硬币观察其正面与反面出现的情况. 其样本空间由四个样本点组成，即 $\Omega = \{$ (正，正), (正，反), (反，正), (反，反)$\}$. 这里，比如样本点 $\omega =$ (正，反) 表示第一枚硬币抛出正面而第二枚抛得反面.

例 1.2 记录某电话机在一小时内呼唤次数，其样本点是非负整数. 当然，我们可以根据实际情况为电话呼唤次数确定一个上界 N. 但是今后会看到，人们宁愿不确定这个 N，而假设呼唤次数可以是任何非负整数，即认为样本空间 $\Omega = \{0, 1, 2, \cdots\}$.

例 1.3 接连射击直到命中为止. 为了简洁地写出其样本空间，我们约定以 "0" 表示一次射击未中，而以 "1" 表示命中. 则样本空间 $\Omega = \{1, 01, 001, \cdots\}$，它是一个可列的无穷集合.

例 1.4 接连不断地射击下去. 沿用上例的记号. 由于这个试验是无休止地反复射击，故其样本点是由 0 与 1 组成的无穷序列. 为确定起见，我们在每个序列之前加上小数点. 于是其样本空间是整数部分为 0 的二进小数全体，即 $\Omega = \{0.a_1a_2a_3\cdots : a_i = 0 \text{ 或 } 1\} = [0, 1]$. 这是一个不可列的无穷集合.

例 1.5 记录某地的最低气温与最高气温. 我们以 x, y 分别表示最低与最高气温，则样本点是数偶 (x, y). 虽然一次试验的结果不能完全预言，但人们总可以确定此时此地气温的下界 a 与上界 b. 于是其样本空间 $\Omega = \{(x, y) : a \leq x \leq y \leq b\}$. 这是坐标平面中一个三角形，包含不可列无穷多个样本点.

不难看出，样本空间 Ω 可以是数集，也可以是任何抽象的集合；可以是有限集，可列集，也可以是不可列的无穷集合；可以是一维的也可以是多维集合. 我们指出，样本空间是研究随机现象的数学模型. 正确地确定不同随机试验的样本点与样本空间是极为重要的. 比如例 1.1 已给出抛两枚硬币观察其正反面出现情况的样本空间，如果我们只关心抛两枚硬币出现正面的个数，则样本空间应改为 $\Omega=\{0,1,2\}$.

1.1.3 事件及其运算

我们时常会关心试验的某一部分可能结果是否出现. 如在例 1.2 中，若以每小时是否达到 5 次电话呼唤来区分这台电话机是否太繁忙，那么"不太繁忙"即不足 5 次的呼唤. 它由样本空间中前 5 个样本点 0, 1, 2, 3, 4 组成. 由于它是由 Ω 中的一部分样本点组成的子集合，故在未来的一次试验中可能发生也可能不发生. 称这种由部分样本点组成的试验结果为**随机事件**，简称作**事件**. 通常用大写的字母 $A, B, \cdots$ 等表示事件. 例如在前述例子中，"不太繁忙"可表为事件 $A=\{$ 至多 4 次呼唤 $\}=\{0,1,2,3,4\}$. 还可以有事件 $B=\{$ 至少 2 次呼唤 $\}=\{2,3,4,\cdots\}$，等等. 于是，如果试验的结果是记录到 1 次呼唤，即样本点 $\omega=1$ 出现，此时事件 A 的要求满足，$\omega\in A$，即事件 A 已发生；事件 B 的要求没满足，$\omega\notin B$，即事件 B 没发生.

我们指出，当样本空间为有穷或至多可列无穷的集合时，可取其任何子集为事件. 而当样本空间为不可列无穷时，比如对例 1.4 中的 $\Omega=[0,1]$，则只能取 Ω 的一部分性质较好的 (称作可测的) 子集作为随机事件. 这一点将在后面的 1.4.2 节作进一步讨论. 由于样本空间 Ω 包含了全部可能结果，因此在每次试验中 Ω 都会发生，故称 Ω 为**必然事件**. 相反，空集 $\varnothing$ 不含任何样本点，每次试验必定不发生，故称 $\varnothing$ 为**不可能事件**. 除此之外，每一随机试验都含有许多随机事件. 由于它们共处于同一试验之中，因而是相互联系着的. 我们有必要弄清它们之间的关系，并引进事件间的运算，以便化复杂事件为简单事件，更好地解决相应的概率问题.

表 1.2 测度论与概率论概念的对照

记　　号	测度论含义	概　率　论　含　义
Ω	空间或全集	样本空间或必然事件
$\varnothing$	空集	不可能事件
ω	元素	样本点
A	可测子集	随机事件
$\omega \in A$	ω 是 A 的元	事件 A 发生
$A \subset B$	A 包含在 B 中	A 发生则 B 必发生
$A = B$	A 与 B 相等	A 与 B 同发生或同不发生
$A \bigcap B$ 或 AB	交集	A 与 B 都发生
$A \bigcap B = \varnothing$	不相交	A 与 B 不相容 (互斥)
$A \bigcup B$	并集	A 与 B 至少一个发生
$\overline{A}$	余集	A 不发生 (对立事件)
$A - B$	差集	A 发生但 B 不发生
$\overline{\lim\limits_n} A_n = \bigcap\limits_{k=1}^{\infty} \bigcup\limits_{n=k}^{\infty} A_n$	上限集	$\{A_n\}$ 中有无限多个发生
$\underline{\lim\limits_n} A_n = \bigcup\limits_{k=1}^{\infty} \bigcap\limits_{n=k}^{\infty} A_n$	下限集	$\{A_n\}$ 中至多有限个不发生
$P(A)$	集 A 的测度	事件 A 的概率
$\xi(\omega)$	可测函数	随机变量
$E(\xi) = \int_\Omega \xi dP$	积分	数学期望

注意我们已经开始把概率论的基本概念纳入测度论轨道: 样本空间就是集论中的空间 (即全集), 样本点是空间的元素, 随机事件就是可测子集, 出现的样本点 ω 是集合 A 的元素则意味着事件 A 发生. 循此可沿用测度论的语言完整地叙述概率论的基本概念, 我们将在后面详述. 这里先将其基本概念间的对照简要地列在表 1.2 中, 请读者结合具体例子进一步理解与熟练运用这些记号与概念.

与集合运算类似, 事件的并与交运算也可推广到任意多个事件的情形. 例如 $\bigcap\limits_{n=1}^{\infty} A_n$ 表示事件列中每一个 A_n 均发生, 而 $\bigcup\limits_{n=1}^{\infty} A_n$ 则表示此事件列中至少有一个 A_n 发生.

可以用图示的方法表示事件之间的关系与运算. 设想向正方形 Ω 内任投一点 ω, 则 Ω 与 ω 分别是此试验的样本空间与样本点, 而正方形 Ω 内的子集 A 代表随机事件. 所投的点 ω 落入 A 中, 则事件 A 发生. 于是事件间的几种关系与运算可表示如图 1.2. 这种表示方法称为维恩图.

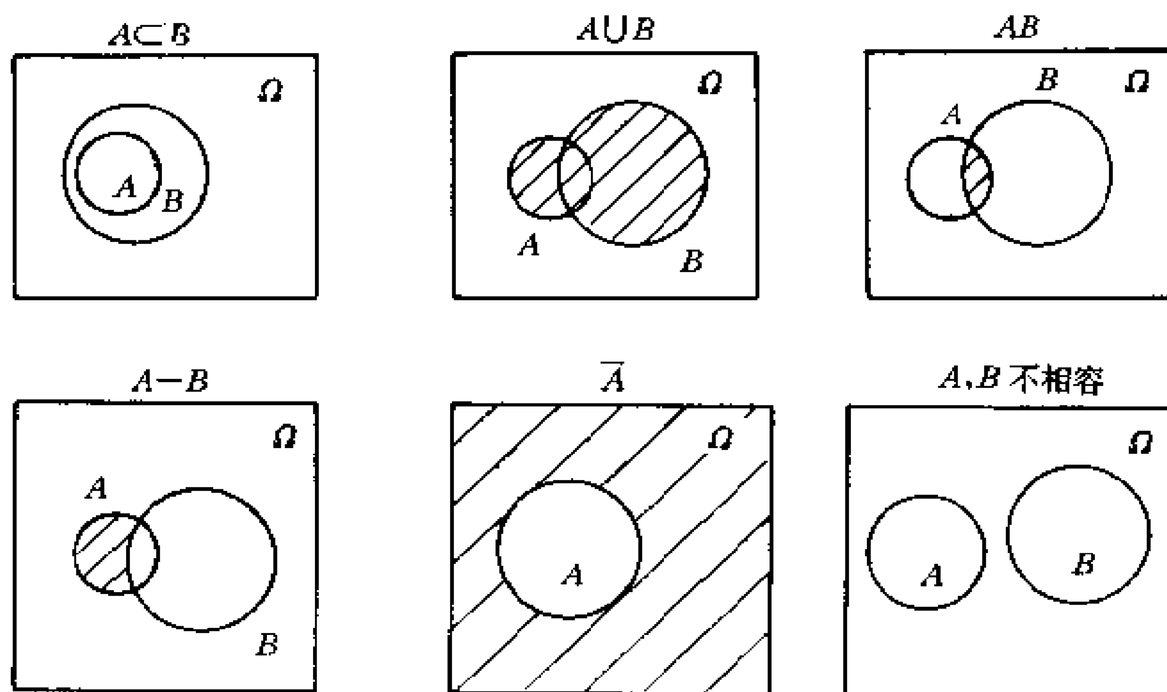


图 1.2　事件运算的维恩图

事件运算有与代数运算类似的交换律、结合律与分配律, 也有作为集合运算的特殊的运算律. 例如有以下对偶律:

$$\overline{A\bigcup B}=\overline{A}\bigcap\overline{B},\qquad \overline{A\bigcap B}=\overline{A}\bigcup\overline{B}. \tag{1.1}$$

我们用概率论语言证明 (1.1) 中第一个等式. 为此需证明“若事件 $\overline{A\bigcup B}$ 发生, 则 $\overline{A}\bigcap\overline{B}$ 发生; 反之, 若 $\overline{A}\bigcap\overline{B}$ 发生则 $\overline{A\bigcup B}$ 也发生”. 事实上若事件 $\overline{A\bigcup B}$ 发生, 则事件“A 与 B 至少一个发生”不发生, 则 A 与 B 全不发生, $\overline{A}$ 与 $\overline{B}$ 全发生, 此即事件 $\overline{A}\bigcap\overline{B}$ 发生. 类似可证“若 $\overline{A}\bigcap\overline{B}$ 发生则 $\overline{A\bigcup B}$ 也发生”. 至此 (1.1) 中第一个等式得证. 对偶律对于多个事件的交、并运算仍然成立.

1.1.4 频率与概率

本节将通过频率引进事件发生的概率，即介绍概率的"统计定义". 在一定条件下，将一随机试验重复 n 次，如果其中事件 A 共发生 m 次，则称

$$F(A) = \frac{m}{n}$$

为事件 A 发生的**频率**.

这里试验次数 n 是自然数，A 发生的次数 m 是非负整数. 因而任何事件 A 的频率 $F(A)$ 总是非负的. 其次，由于必然事件 Ω 在每次试验中均发生，故其频率 $F(\Omega) = n/n = 1$. 最后，当事件 A 与 B 互不相容，即 A 与 B 不可能同时发生时，事件 $A \bigcup B$ 发生的次数等于 A 发生的次数与 B 发生的次数之和. 于是频率有可加性，即事件 $A \bigcup B$ 发生的频率等于 A, B 各自发生频率之和. 总之，频率 $F(A)$ 有如下基本性质：

1° **非负性**：$F(A) \geq 0$;

2° **规范性**：$F(\Omega) = 1$;

3° **可加性**：A 与 B 不相容时 $F(A \bigcup B) = F(A) + F(B)$.

前已指出，事件发生的频率有稳定性，即当重复试验的次数 n 很大时，每个事件 A 发生的频率 $F(A)$ 有一个稳定值. 例如，蒲丰等人的试验结果表明，大量重复地抛一枚均匀硬币时，正面出现的频率稳定在 0.5 附近. 这是因为硬币的质地与形状都是均匀的，正面与反面出现的可能性相等. 于是频率的稳定值 0.5 恰好代表了正面出现可能性的大小. 又如，统计了大量英文文献的字母使用情况，发现字母 E 使用的频率稳定在 0.105 附近，而字母 J 使用的频率则小得多，大约为 0.001. 显然，这种字母使用频率的大小取决于英语本身，由它的特定的构词方法与语法所决定. 频率的稳定值反映了各个字母在英语中出现的可能性的大小.

这种表征在一定条件下事件 A 发生可能性大小的频率稳定值，

就称作事件 A 的概率. 记为 $P(A)$. 作为频率的稳定值, 概率 $P(A)$ 也应具备相应的三条性质:

1° **非负性**: $P(A) \geq 0$;

2° **规范性**: $P(\Omega) = 1$;

3° **可加性**: 若 A 与 B 不相容, 则 $P(A \bigcup B) = P(A) + P(B)$.

我们指出, 在一定条件下进行一随机试验, 各确定事件发生的概率是这个试验固有的性质, 它是由试验的条件与各事件确定的内含所决定的. 可以通过大量地重复此试验, 借助频率的稳定值去认识它, 也可以通过其它途径, 比如根据对试验机制的具体分析来确定. 严格地说, 有些事件的概率是无法通过重复试验来确定的. 例如气象台每晚要发布未来一天中北京市的"降水概率", 这个概率应当是根据采集到的数据经科学的分析与计算得出的. 特定时日的北京市在历史的长河中只有一个, 当然无法大量重复地观测其降水情况, 从而也无法借助频率的稳定值求出降水概率.

总之, 我们借助大量重复试验中频率值呈现的稳定性, 来说明表征事件发生可能性大小的概率是客观存在的. 这里所使用的"大量试验中频率的稳定值", 是一种极不规范的说法. 实际上, 概率 $P(A)$ 就是当试验次数 $n \to \infty$ 时, 频率 $F(A)$ 的一种极限. 只不过其中 A 发生的频数 m 是不能完全预言的数量, 从而 $F(A) = m/n$ 也不是通常的数列而已. 其严格的含义将在第四章详加叙述.

1.1.5 习题

1. 设 A, B, C, D 是某一试验中的四个事件. 试用它们的运算表达如下事件: (1) 四个事件中至少一个发生; (2) 恰好发生两个; (3) 至少发生三个; (4) 至多发生一个.

2. 设 A, B, C 为某随机试验中三个事件. 试说明下列关系式的概率含义, 并画出相应的维恩图:

(1) $A \bigcap B \bigcap C = A$; (2) $A \bigcup B \bigcup C = A$;

(3) $A \bigcap B \subset C$; (4) $A \subset \overline{B \bigcap C}$.

3. 求证事件运算的分配律:

$(A\bigcup B)\bigcap C=(A\bigcap C)\bigcup(B\bigcap C)$,

$(A\bigcap B)\bigcup C=(A\bigcup C)\bigcap(B\bigcup C)$.

4. 记录某电话机在一小时内呼唤次数. 设 $A_k=\{$ 至少 k 次呼唤 $\}$, 试述事件 $\overline{A}_k, A_k-A_{k+1}, \bigcup\limits_{k=1}^{\infty}A_k, \bigcap\limits_{k=0}^{\infty}A_k$ 的含义.

5. 设事件列 $\{A_n\}$ 单调上升: 任 $n\geq 1$ 有 $A_n\subset A_{n+1}$. 试用概率论语言证明

$$\overline{\lim_{n}}A_n=\bigcup_{n=1}^{\infty}A_n=\underline{\lim_{n}}A_n.$$

对单调下降事件列, 相应的结论是什么?

6. 写出如下各试验的样本空间及指定事件所含样本点.

(1) 将一枚硬币抛三次, $A=\{$ 第一次为正面 $\}$, $B=\{$ 三次出现同一面 $\}$, $C=\{$ 有正面 $\}$.

(2) 将一枚骰子掷两次, $A=\{$ 点数相同 $\}$, $B=\{$ 其中一枚点数是另一枚的 2 倍 $\}$, $C=\{$ 点数之和为 6$\}$.

(3) 将一红球与一黑球任放入编号为 1,2,3 的三个盒中, $A=\{1$ 号盒不空 $\}$, $B=\{1$ 号盒与 2 号盒各一球 $\}$, $C=\{$ 每盒至多一球 $\}$.

7. 试画出表示 6 题中各事件间关系的维恩图.

8. 从 0 至 9 这 10 个数字中有重复地取 5 个. 记 E_i 为某些数字被取出 i 次 (例如取出 52353, 则 E_0, E_1, E_2 三事件均发生). 试画出表示 $E_0, E_1, E_2, E_3, E_4, E_5, E_6$ 之间关系的维恩图.

1.2 古典概型

1.2.1 模型的定义

最先涉及到的求概率问题都满足"各可能结果具有等可能性"这一假设. 例如, 在游戏中使用的骰子是一个匀质的正方体, 使得掷出 1 至 6 各个点数的可能性相同, 以保证游戏的公平. 又如一副纸牌的每一张形状与大小均相同, 而且在发牌前还需充分地洗牌, 于是发到其中每张牌也是等可能的.

定义 2.1 如果一随机试验具有以下特点:

(1) 样本空间只含有限多个样本点, 即 $\Omega = \{\omega_1, \cdots, \omega_n\}$;

(2) 各样本点出现的可能性相等, 即对每个 $i = 1, \cdots, n$ 有

$$P(\{\omega_i\}) = \frac{1}{n}, \tag{2.1}$$

则称此随机试验为**古典型** 的. 此时对每一事件 $A \subset \Omega$, 可取

$$P(A) = \frac{n(A)}{n(\Omega)}, \tag{2.2}$$

称此 $P(A)$ 为事件 A 的**古典概率**.

注意, 我们总用 $n(A)$ 表示有限集 A 中元素的个数. 易见这样定义的 P 满足上节的非负性与规范性, 再由不交集之并的计数等于各集计数之和, 即当 $A \bigcap B = \emptyset$ 时有 $n(A \bigcup B) = n(A) + n(B)$, 立得 P 具有可加性.

至此, 当我们确信一随机试验是古典型的, 即定义中条件 (1) 与 (2) 满足之后, 便可运用 (2.2) 式计算此试验中各事件发生的概率. 于是, 问题归结为数一数样本点总数 $n(\Omega)$ 与所涉及的事件 A 中有利场合数 $n(A)$. 由于计数过程有时也相当复杂, 在此有必要简述初等数学中的计数原理.

1.2.2 计数原理

按计数过程的串行与并行, 我们有如下两条计数原理:

乘法原理: 假定进行过程 I 有 n_1 种方式, 而对于过程 I 的每一方式, 进行过程 II 都有 n_2 种方式. 那么, 依次进行过程 I 与 II 共有 $n_1 n_2$ 种方式.

加法原理: 假定进行过程 I 有 n_1 种方式, 进行过程 II 有 n_2 种方式. 那么, 进行过程 I 或 II 共有 $n_1 + n_2$ 种方式.

上述两个原理都可以推广到多个过程的场合.

运用乘法原理立即可得如下结论:

从 n 个不同的元素中，有放回地取 r 个元素组成的**可重复排列**的种数为 $nn\cdots n=n^r$. 从 n 个不同的元素中，不放回地取 $r(r\leq n)$ 个组成的**不重复排列** 的种数为 $A_n^r=n(n-1)\cdots(n-r+1)$. 从 n 个不同的元素中，不放回地取 $r(r\leq n)$ 个组成的**组合** 种数为

$$\binom{n}{r}=\frac{n(n-1)\cdots(n-r+1)}{r!}=\frac{n!}{r!(n-r)!}.$$

注意取样分"有放回"与"不放回"两种方式. 如果每次取出一个，记下其特征后放回，经充分混合后再取下一个，这种取样方式叫做"有放回"，所得样品是有重复的. 如果每次取出一个，记下其特征后不放回，再从余下的对象中取下一个，这种取样方式为"不放回"，所得样品是不重复的.

定理 2.1 组合有如下性质:

1° $\binom{n}{r}=\binom{n}{n-r}$;

2° $\binom{n}{r}=\binom{n-1}{r-1}+\binom{n-1}{r}$;

3° $\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$;

4° $\binom{n}{1}+2\binom{n}{2}+\cdots+n\binom{n}{n}=n2^{n-1}$;

5° $\sum\limits_{i=0}^{k}\binom{n}{i}\binom{m}{k-i}=\binom{n+m}{k}$;

6° $\binom{n}{0}^2+\binom{n}{1}^2+\cdots+\binom{n}{n}^2=\binom{2n}{n}$.

可以直接将组合数的定义代入，或者运用其它分析技巧证明上述性质，也可以通过设计特定的计数过程来证明它们. 例如为证性质 2°，我们把从 n 个元素中任取 r 个的组合，分为两个并行的计数过程: 含 1 号元素 (过程 I) 与不含 1 号元素 (过程 II). 它们分别有 $\binom{n-1}{r-1}$

与 $\binom{n-1}{r}$ 种方式，故由加法原理得证性质 2° 成立．为熟练掌握计数原理，我们提倡读者用这种方法完成定理的证明．

下面考虑组合概念的推广．从 n 个元素中任取 r 个的组合，可理解为将这 n 个元素任意分为两堆：第 1 堆 $r_1 = r$ 个取走，第 2 堆 $r_2 = n - r$ 个留下．于是，这个组合数可改用以下记号

$$\binom{n}{r_1\ r_2} = \frac{n!}{r_1!r_2!}. \tag{2.3}$$

循此可把组合推广到分 n 个元素为 k 堆的一般场合，其中第 1 堆含 r_1 个元素，第 2 堆含 r_2 个，$\cdots$，第 k 堆含 r_k 个，$r_1+r_2+\cdots+r_k = n$．那么，反复运用 (2.3) 式及乘法原理可得，上述分为 k 堆的分法共有

$$\binom{n}{r_1\ r_2\ \cdots\ r_k} = \frac{n!}{r_1!r_2!\cdots r_k!} \tag{2.4}$$

种．

组合数 $\binom{n}{r}$ 还可朝另一个方向推广：保持 r 为非负整数不变，将自然数 n 换为任意的实数 α，定义

$$\binom{\alpha}{r} = \frac{\alpha(\alpha-1)\cdots(\alpha-r+1)}{r!}. \tag{2.5}$$

由定义立得，当 α 为小于 r 的整数时 $\binom{\alpha}{r} = 0$.

最后强调指出，当我们运用计数原理完成较复杂的计数任务时，适当地分解计数过程 (串行或者并行) 是十分重要的．为了便于迅速又正确地计数，可以人为地分解为若干计数过程，而与我们的自然计数过程不尽相同．这一点将通过具体例子作进一步的说明．

至此我们已经作好了计算古典概率的准备工作．

1.2.3 例题

例 2.1 抛两枚均匀硬币，求出现正面反面各一个的概率．

解: 上节例 1.2 已列出此试验的样本空间 $\Omega = \{$(正, 正), (正, 反), (反, 正), (反, 反)$\}$. 由硬币的均匀性知 4 个样本点是等可能的, 所以是古典型. 事件 $A = \{$ 正面反面各一个 $\} = \{$(正, 反), (反, 正)$\}$ 含两个样本点. 故由 (2.2) 式知所求概率 $P(A) = 1/2$.

历史上曾有人认为此题的解是 1/3. 理由是样本空间由 $\omega_1 = \{$ 两个正面 $\}$, $\omega_2 = \{$ 一个正面一个反面 $\}$, $\omega_3 = \{$ 两个反面 $\}$ 这三个样本点组成. 而事件 A 只含其中 ω_2 这一个样本点, 故 $P(A) = 1/3$. 此解法的错误在于, 这样取样本空间 $\Omega' = \{\omega_1, \omega_2, \omega_3\}$, 三个样本点不是等可能的, 即 Ω' 不是古典型样本空间 (今后会看到, 这是所谓的贝努里概型), 当然不能用古典概率的计算公式 (2.2).

例 2.2 取数问题 从 0 至 9 这十个数字中不放回地任取 4 个排好, 求恰排成一个 4 位 (即 1000 以上) 偶数的概率.

解: 样本点是不重复排列, 其总数 $n(\Omega) = A_{10}^4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$. 由任意性知各样本点等可能, 故属于古典概型. 只需再确定事件 $E = \{$ 恰排成 4 位偶数 $\}$ 中样本点的个数. 考虑到偶数的末位必为 0,2,4,6,8 中的某一个, 而按题意 0 又不能作千位数. 于是将对事件 E 中样本点的计数, 分解为 "末位是 0" 与 "末位是 2,4,6,8 中某一个" 这两个并行的过程. 据乘法原理, 前一过程有 $9 \cdot 8 \cdot 7 \cdot 1 = 504$ 种方式; 后一过程有 $8 \cdot 8 \cdot 7 \cdot 4 = 1792$ 种方式. 再由加法原理便得 $n(E) = 504 + 1792 = 2296$. 代入 (2.2) 可得所求概率 $P(E) = 0.46$.

从以上两个简单例子可以看出, 解古典型概率问题的关键步骤是

(1) 确定该随机试验的样本点与样本空间;

(2) 确认其样本空间是由有限个等可能的样本点组成;

(3) 数清样本点总数与所关心事件中样本点个数, 代入 (2.2) 式算出概率.

其中步骤 (1) 是正确解决问题的基础. 初学者经常出现的错误有: 取错了样本点; 解同一问题时样本点取法不一致; 样本点选取不明确导致计数过程中出现重复、遗漏等. 这些都需要用心注意避免.

为叙述简明，我们在介绍较复杂的例题时，未必把上述步骤一一写出.

例 2.3 占位问题 r 个不同的球任意放入编号为 1 至 n 的 n 个盒 $(r \le n)$，每球入各盒均等可能. 试求下列事件的概率：

$A = \{$ 指定 r 个盒恰各含 1 球 $\}$； $B = \{$ 每盒至多 1 球 $\}$；

$C = \{$ 某指定盒恰含 m 个球 $\}$.

解 样本点是从盒号 1 至 n 中任选 r 个的有重复的排列. 总数 $n(\Omega) = n^r$. 事件 A 中样本点取决于 r 个球放在 r 个盒中的顺序，故 $n(A) = r!$，从而有 $P(A) = r!/n^r$. 事件 B 与事件 A 的差异仅在于各含 1 球的 r 个盒没有指定，用乘法原理可得到 $n(B) = \binom{n}{r} \cdot n(A) = \binom{n}{r} r!$，故 $P(B) = \binom{n}{r} r!/n^r$. 为数清事件 C 所含样本点数 $n(C)$，我们先取 m 个球放入指定盒中，共有 $\binom{r}{m}$ 种取法. 然后再把剩下的 $(r-m)$ 个球任意放在其余 $(n-1)$ 个盒中，放法有 $(n-1)^{r-m}$ 种. 据乘法原理可得 $n(C) = \binom{r}{m}(n-1)^{r-m}$. 故 $P(C) = \binom{r}{m}(n-1)^{r-m}/n^r$.

占位问题在统计物理学中有重要的应用. 这里的“球”可理解为粒子，而“盒”则代表粒子的相位. 我们介绍的是“诸球可辨”(不同)且“各盒内球数不限”的模型. 此外尚有“球不可辨”“每盒至多一球”等其它模型，在此不一一详述.

占位问题的一个特例是如下的

例 2.4 生日问题 求任意 r 个人生日各不相同的概率.

解 理解一年 365 天就是 365 个盒. r 个人去任意占这 365 个盒. “生日互不相同”即每盒至多一球. 此即上例的事件 B. 于是所求概率为

$$P(B) = \frac{\binom{365}{r} r!}{365^r} = \frac{365 \cdot 364 \cdots (365 - r + 1)}{365^r}.$$

当 $r=30$ 时，$P(B)=0.294$, 而当 $r=55$ 时，$P(B)=0.01$. 就是说，任意 55 个人，至少有两人生日相同的概率大到 99%, 几乎等于必然事件的概率了. 我们知道，任意 366 个人至少二人生日相同是必然事件，而 55 离 366 相差甚远，却得到了几乎一样的结论，这是我们直接经验所意想不到的.

例 2.5 公平的抽签 袋中有 r 个红球与 b 个黑球，现任意不放回地一一摸出，求事件 $R_k=\{$ 第 k 次摸出红球 $\}$ 的概率.

解 1: 视同色球是可辨的，可理解球不仅着了色，还编上不同的号码. 于是样本点是这 $(r+b)$ 个球的全排列, 即 $n(\Omega)=(r+b)!$. 摸球的任意性保证各样本点等可能, 故是古典型. 为数清事件 R_k 中样本点数，我们分解它为两个串行的过程：先从 r 个红球中选一个放在第 k 个位置上. 共有 r 种选法. 再将其余 $(r+b-1)$ 个球在其余的位置上任意排列, 有 $(r-b+1)!$ 种方式. 由乘法原理立得 $n(R_k)=r(r+b-1)!$, 故得 $P(R_k)=r/(r+b)$.

解 2. 视同色球不可辨. 此时样本点取决于这 $(r+b)$ 个位置中哪 r 个放红球，共有 $n(\Omega)=\begin{pmatrix} r+b \\ r \end{pmatrix}$ 个样本点. 而 R_k 要求第 k 个位置必须是红球, 于是只需从 $(r+b-1)$ 个位置中选 $(r-1)$ 个放其余的红球，即 $n(R_k)=\begin{pmatrix} r+b-1 \\ r-1 \end{pmatrix}$. 代入可算得 $P(R_k)=r/(r+b)$.

两种解法所得结果相同, 都与取球次序 k 无关. 这证明抽签是公平的: 如果 $(r+b)$ 个人将有 r 个人中签, 那么无论是先抽还是后抽, 即无论 $k=1,2,\cdots,r+b$, 每人中签的概率均为 $P(R_k)=r/(r+b)$, 与抽取次序无关.

例 2.6 抽样检验 袋中有 r 个红球与 b 个黑球. 现从中任取 n 个，试对 (1) 有放回；(2) 不放回 二方式求事件 $E=\{n$ 个中恰含 k 个红球 $\}$ 的概率.

解:　**不放回情形**

样本点是组合，$n(\Omega)=\binom{r+b}{n}$. 由乘法原理可得 $n(E)=\binom{r}{k}\binom{b}{n-k}$. 故

$$P(E)=\frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}},\quad r=0,1,\cdots,n. \tag{2.6}$$

有人认为 $n(E)=\binom{r}{k}$，可能是因为只从字面上注意到"恰含 k 个红球"的缘故. 实际上是错把选出的 k 个红球当作样本点，而不是 Ω 中由 n 个球组成的样本点，就是说样本点含义不一致了. "恰含 k 个红球"当然意味着"其余 $(n-k)$ 个均为黑球"，这是绝不可遗漏的.

放回情形

样本点是可重复的排列，样本点总数 $n(\Omega)=(r+b)^n$. 为数清 E 中所含样本点数，我们分解为三个串行的过程：先确定这 n 个球中哪 k 个位置上是红球，共有 $\binom{n}{k}$ 中方式；再从 r 个红球中有重复地选取 k 个，有 r^k 种取法；最后，从 b 个黑球中有重复地取 $(n-k)$ 个，有 b^{n-k} 种取法. 用乘法原理知 $n(E)=\binom{n}{k}r^kb^{n-k}$. 故有

$$P(E)=\binom{n}{k}\left(\frac{r}{r+b}\right)^k\left(\frac{b}{r+b}\right)^{n-k},\ k=0,\cdots,n. \tag{2.7}$$

这个例子可用于产品的抽样检验. 以红球代表正品，黑球代表次品. 则 $P(E)$ 就是任取 n 个样品中恰含 k 个正品的概率. 两种取样方式所得到的概率 (2.6) 与 (2.7) 式，分别称作超几何分布与二项分布，将在第二章进一步研究.

例 2.7 结绳问题 将 n 根绳的 $2n$ 个头任意两两相接，试求事件 $A=\{$ 恰结成 n 个圈 $\}$ 与 $B=\{$ 恰结成一个圈 $\}$ 的概率.

解： 将 $2n$ 个头任意排成一列，然后使第 1 个头与第 2 个头相接，第 3 个头与第 4 个头相接，循此下去，最后第 $(2n-1)$ 个头与

第 $2n$ 个头相接. 因此, 一种接法相当于 $2n$ 个头的一个全排列. 故 $n(\Omega)=(2n)!$.

为实现事件 A 的要求, 则每根绳的头必须与自己的尾相接, 即每根绳的两个头必须排在第 $(2k-1)$ 与第 $2k$ 个位子上 $(k=1,2,\cdots,n)$. 于是 n 根绳的位置共有 $n!$ 种排法. 又对应于上述每一种排法, 各根绳的头与尾又都有两个排法, 根据乘法原理可得 $n(A)=n!2^n=(2n)!!$. 于是得到

$$P(A)=\frac{(2n)!!}{(2n)!}=\frac{1}{(2n-1)!!}.$$

下面再求事件 B 中样本点个数. 不妨直接数一数满足 "恰成一个圈" 要求的 $2n$ 个头的排列种数. 位于第 1 位的头有 $2n$ 种选法, 然后, 第 2 位的头共有 $(2n-2)$ 种选法, 其中排除了已选定位于第 1 位的绳的头与尾. 将已选好的两个头相接, 就剩下 $(n-1)$ 根绳的 $(2n-2)$ 个头. 于是, 选择第 3 个头有 $(2n-2)$ 种方法, 再选取第 4 个头有 $(2n-4)$ 种方法, $\cdots$ 循此下去, 用乘法原理便得 $n(B)=2n(2n-2)(2n-2)(2n-4)\cdots4\cdot2\cdot2\cdot1=(2n)!!(2n-2)!!$. 从而我们得到

$$P(B)=\frac{(2n)!!(2n-2)!!}{(2n)!}=\frac{(2n-2)!!}{(2n-1)!!}.$$

古典概率的例题很多, 所用方法灵活多样. 利用定义直接计算古典概率的例子先介绍到这里.

1.2.4 习题

1. 设计适当的试验证明定理 2.1 中组合性质 3° 至 6°.

2. 设计适当的试验, 求不定方程 $x_1+x_2+\cdots+x_n=r$ 的 (1) 非负整数解的个数; (2) 正整数解的个数.

3. 袋中有 r 个红球与 b 个黑球, 现一一摸出, 直至只剩下同色球为止, 求剩下的全是红球的概率.

4. 袋中有 r 个红球与 b 个黑球, 现有放回地任意摸球, 直到摸出 i 个红球为止, 求恰好摸 k 次的概率.

5. 从 n 对不同的鞋子中任取出 $2m$ 只 $(2m < n)$, 求其中恰有 k 双配对的鞋子的概率.

6. 甲抛 $(n+1)$ 个, 乙抛 n 个均匀硬币, 求甲所得正面比乙多的概率.

7. 某公共汽车载有 8 名乘客自始点发车, 此后共停 10 个车站而到达终点. 假定每个乘客在各站下车均为等可能. 求下列事件的概率: (1) $A=\{$ 每站至多 1 人下车 $\}$; (2) $B=\{$8 个人在同一站下车 $\}$; (3) $C=\{$8 个人都在最后一站下车 $\}$; (4) $D=\{$8 个人中恰有 3 个人在最后一站下车 $\}$; (5) $E=\{$ 最后一站已无人下车 $\}$.

8. 将 n 个不同的球任放入 n 个不同的盒, 每球入各个盒均为等可能. 试求 (1) 无空盒; (2) 1 号盒空着的概率.

9. 将 r 个相同的球任放入编号为 1 至 n 的 n 个盒 $(r \leq n)$, 每球入各盒均等可能. 试求 $A=\{$ 指定 r 个盒恰各 1 球 $\}$; $B=\{$ 每盒至多 1 球 $\}$; $C=\{$ 指定盒恰有 m 个球 $\}$ 的概率.

10. 将 3 个不同的球任放入编号为 1,2,3,4 的四个盒, 每球入各盒均为等可能. 求 (1) 恰两个空盒; (2) 有球盒的最小编号为 2 的概率.

11. 从一副 52 张扑克牌中任取出 5 张, 求下列事件的概率: (1) 以 A 打头的同花顺次 5 张; (2) 其它同花顺次 5 张; (3) 有 4 张同点数; (4) 3 张同点数另外 2 张也同点数; (5) 5 张同花; (6) 异花顺次 5 张; (7) 3 张同点另 2 张不同点; (8) 5 张中有两对; (9) 5 张中恰有 1 对; (10) 其它情形.

12. 将一副 54 张扑克牌任意等分给 6 个人, 试求双王及 4 个 A 都分给某一个人的概率.

13. 将 r 个红球与 b 个黑球任排成一列 $(r \leq b)$, 求没有两个红球相邻的概率.

14. 袋中装有编号为 1 至 N 的 N 个球, 现有放回地任取 n 个, 依次记下其号码, 求 (1) n 个号码按严格上升排列; (2) n 个号码非降排列的概率.

15. 有 2^n 名选手 (含甲与乙) 参加一场淘汰赛. 第一轮将他们任意两两配对比赛, 然后 2^{n-1} 名胜者再任意两两配对进行第二轮比赛, 循此下去, 直至第 n 轮决出一名冠军为止. 假定每一名参赛选手在各轮比赛中胜负均是等可能的. 求甲与乙在这场比赛中相遇的概率.

1.3 几何概型

1.3.1 定义及例

本节继续考查样本点的出现有等可能性的随机试验. 但不是象古典概型那样局限于有限多个样本点的情形. 有时, 虽然试验的可能结果是无限多的, 但由于试验的任意性或对称性, 各样本点出现的机会仍然相等. 当然, 此时不能简单地通过样本点的计数来计算概率.

假定在盛有 1 升水的容器中有一个任意游动的细菌. 现从容器的任意位置用吸管取出 10 毫升的水样, 由于细菌运动与取水样的任意性, 10 毫升水样中含有这个细菌的概率应为百分之一, 即等于水样的体积与水的总体积之比.

象这样的例子还有很多, 它们的共同特点是通过空间集合的几何度量 (体积, 面积, 长度等) 来计算概率.

假定试验的可能结果是空间中的点 (如上例中取水时细菌的位置), 所有样本点的集合 Ω 是空间中一个几何图形 (上例中装在容器内的 1 升水). 它可以是 1 维, 2 维, 3 维, 甚至是 n 维的. 样本空间 Ω 及作为随机事件的子集 A 都有有限的几何度量, 即 1 维情形的长度, 2 维情形的面积, 3 维情形的体积等, 统称为勒贝格测度. 由于试验的对称性, 各种结果出现是等可能的, 体现在样本点于 Ω 中均匀分布: 样本点落在 Ω 的每个子集 A 中的概率只与 A 的勒贝格测度 $m(A)$ 成正比, 而与 A 的位置及形状无关. 由于此时 Ω 较为复杂, 作为事件的 Ω 的子集不能太任意. 至少应当是可以度量其大小的, 即 Ω 的可测子集 (参见下节的波雷尔集).

定义 3.1 设 Ω 为 n 维欧氏空间中的确定的集合，满足条件 $0 < m(\Omega) < +\infty$. 对 Ω 的任何可测子集 A, 称

$$P(A) = \frac{m(A)}{m(\Omega)} \tag{3.1}$$

为事件 A 的**几何概率**.

由勒贝格测度的性质立得，(3.1) 式定义的 P 满足概率所应当具备的

1° **非负性**：$P(A) \geq 0$;

2° **规范性**：$P(\Omega) = 1$;

3° **可列可加性**：对任何两两不相容的事件列 $\{A_n\}$ 有

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

下面的 1.4.3 节将证明，可列可加性蕴含 1.1.4 节所述的可加性. 由于这里的 Ω 含无限多个样本点，故有必要考虑涉及到无限多个事件的概率. 以下是几何概型的基本例题.

例 3.1 约会问题 甲乙二人约定在 $[0, T]$ 时段内去某地会面，规定先到者等候一段时间 $t(t \leq T)$ 再离去. 试求事件 $A = \{$ 甲乙将会面 $\}$ 的概率.

解：分别以 x, y 表示甲乙到达会面地点的时间，则样本点是坐标平面上一个点 (x, y)，而样本空间 $\Omega = \{(x, y) : 0 \leq x, y \leq T\}$ 是边长为 T 的正方形. 由于二人到达时刻的任意性，样本点在 Ω 中均匀分布，属几何概型. 我们关心的事件是

$A = \{$甲乙将会面$\} = \{(x, y) \in \Omega : |x - y| \leq t\}$.

如图 3.1, A 是正方形 Ω 中夹于直线 $x - y = t$ 与直线 $x - y = -t$ 中间的部分. 用 (3.1) 式便得所求概率

$$P(A) = \frac{m(A)}{m(\Omega)} = \frac{T^2 - (T - t)^2}{T^2} = 1 - \left(1 - \frac{t}{T}\right)^2.$$

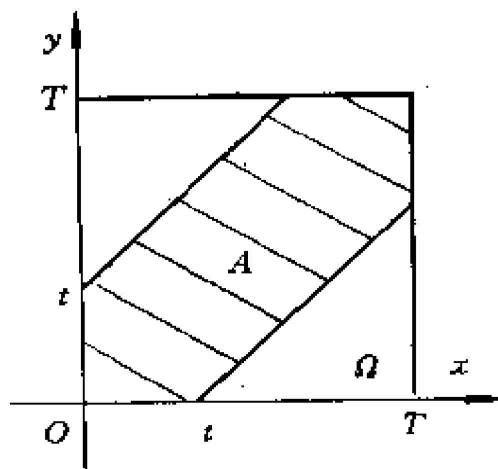


图 3.1 约会问题的模型

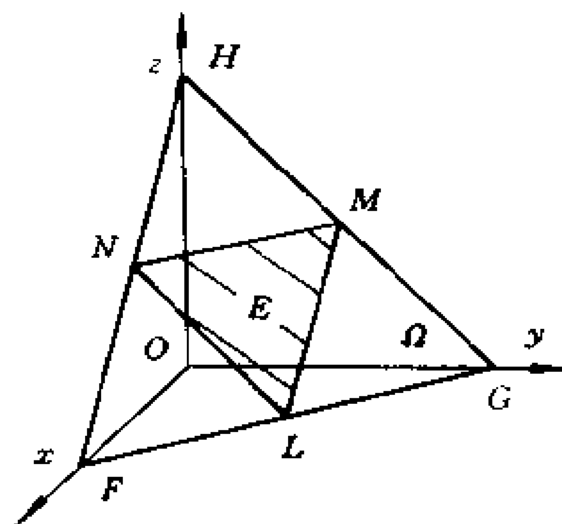


图 3.2 例 3.2 的模型

例 3.2 在圆周上任取三点 A, B, C, 求事件 $E = \{\triangle ABC$ 成锐角三角形 $\}$ 的概率.

解: 不妨假定点 A, B, C 在圆周上顺时针排列, 分别以 x, y, z 表示 $\overset{\frown}{AB}, \overset{\frown}{BC}, \overset{\frown}{CA}$ 的弧度, 于是样本点是三维空间中的点 (x, y, z), 而样本空间为 $\Omega = \{(x, y, z) : x, y, z \geq 0$ 且 $x+y+z = 2\pi\}$. 如图 3.2, Ω 是空间坐标系下的三角形 FGH. 由任意性知样本点在 Ω 中均匀分布, 即属于几何概型. 我们所关心的事件是 $E = \{\triangle ABC$为锐角三角形$\}$ $= \{(x, y, z) \in \Omega : x, y, z \leq \pi\}$, 即图 3.2 中三角形 LMN(阴影部分). 用 (3.1) 式可得

$$P(E) = \frac{m(E)}{m(\Omega)} = \frac{1}{4}.$$

例 3.3 蒲丰投针问题 桌面上画满间隔均为 a 的平行直线, 现向桌面任意投放一长为 $l(l < a)$ 的针, 求事件 $E = \{$ 针与某直线相交 $\}$ 的概率.

解: 如图 3.3, 针的位置由针的中点到最近直线的距离 ρ 及针与直线所夹锐角 θ 所决定. 于是样本空间 $\Omega = \{(\rho, \theta) : 0 \leq \rho \leq a/2,\ 0 \leq \theta \leq \pi/2\}$, 它是坐标平面中一个矩形. 由投针的任意性, 样

本点 (ρ,θ) 在 Ω 中均匀分布，是几何概型. 而针与某直线相交，当且仅当 $\rho \le \frac{l}{2}\sin\theta$. 即事件

$$E=\{(\rho,\theta)\in\Omega:\rho\le\frac{l}{2}\sin\theta\}.$$

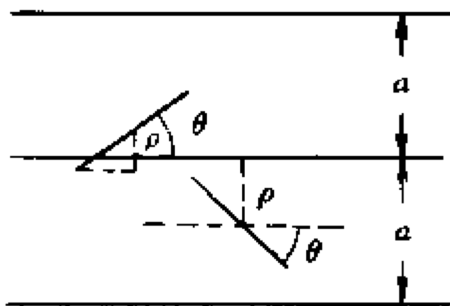


图 3.3 投针问题

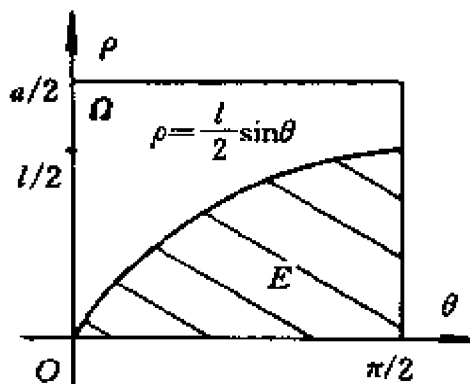


图 3.4 投针问题的模型

图 3.4 画出这一几何概型的图示. 易见

$$m(\Omega)=\frac{\pi a}{4},\quad m(E)=\int_0^{\frac{\pi}{2}}\frac{l}{2}\sin\theta d\theta=\frac{l}{2}.$$

故由 (3.1) 式得

$$P(E)=\frac{m(E)}{m(\Omega)}=\frac{2l}{\pi a}. \tag{3.2}$$

蒙特卡洛方法 将上述投针试验重复 n 次，如果记录到事件 E 发生了 m 次，则由 1.1.4 节可知，当 n 充分大时， E 的频率 m/n 可作为概率 $P(E)$ 的近似值，于是由 (3.2) 式可得

$$\frac{m}{n}\approx\frac{2l}{\pi a},$$

解出

$$\pi\approx\frac{2ln}{am}. \tag{3.3}$$

至此可以通过大量重复投针试验，用 (3.3) 式计算出 π 的近似值. 这个例子的重要性在于，由此开辟了一条通过设计适当的随机试验而完

成某种计算任务的途径，这就是著名的蒙特卡洛方法. 以后将给出运用这一方法的进一步例子.

由以上例题可以看出，几何概型可用于解决几何问题 (如例 3.3 与例 3.2)，也可用于解决非几何问题 (如例 3.1). 但不论哪种情形，总应当画出几何概率的模型图，其上标出样本空间 Ω 及所涉及到的事件. 这个模型图是正确求出几何概率的基础. 上述图 3.1, 图 3.2, 图 3.4 就分别是所举三个例题的模型图.

1.3.2 贝特朗奇论

1889 年，法国数学家贝特朗提出下述几何概率问题，并给出三种不同的答案. 这就使有的人对当时的概率论中的一些概念与方法产生怀疑，因此被称作 "奇论".

例 3.4 在单位圆上任作一弦，求弦长大于 $\sqrt{3}$ 的概率.

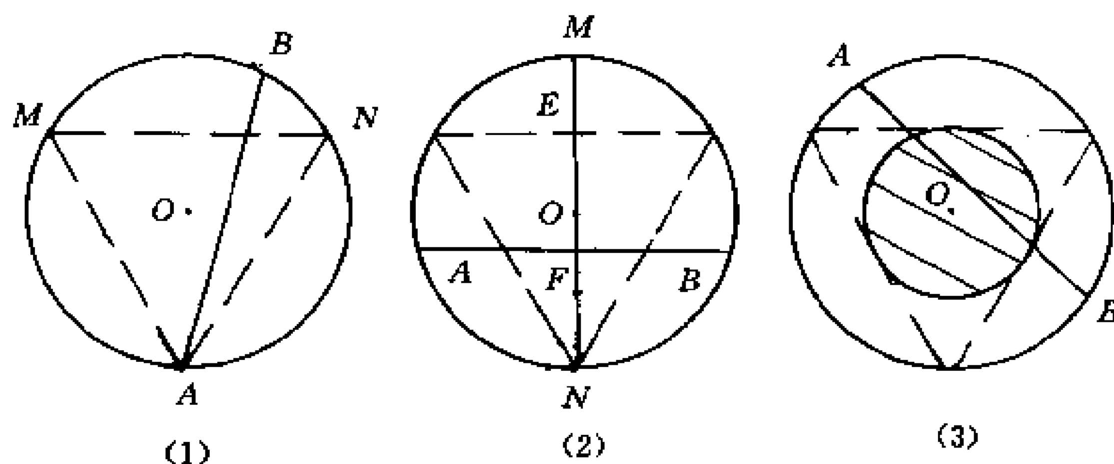


图 3.5 贝特朗奇论

解 1: 如图 3.5 中 (1) 图. 不妨设弦的一端点 A 已取定. 问题化为在圆周上任取另一端点 B. 故样本空间 Ω 为整个圆周. 因为单位圆的内接正三角形 AMN 的边长恰为 $\sqrt{3}$, 故弦长 AB 大于 $\sqrt{3}$, 当且仅当端点 B 落在 $\overset{\frown}{MN}$ 上. 由于 $\overset{\frown}{MN}$ 的长为圆周长的 $1/3$, 故所求概率为 $1/3$.

解 2: 如图 3.5 的 (2) 图. 不妨只考虑与直径 MN 垂直的弦. 当且仅当弦心距小于 $1/2$, 即所作弦的中点在 EF 上时弦长大于 $\sqrt{3}$.

因此所求概率为 1/2.

解 3: 如图 3.5 的 (3) 图. 弦由其中点位置确定. 而弦长大于 $\sqrt{3}$ 的充要条件是弦的中点落在半径为 1/2 的同心圆内. 故所求概率为 1/4.

认真分析上述解题过程可知，产生不同答案的根本原因是题目中"任作一弦"的含义不清. 对"任作"二字持不同的理解，就会得到不同的答案: 理解为在圆周上任取两点连成一弦，则所求概率为 1/3; 理解为在固定直径上任取一点作与此直径垂直的弦，答案为 1/2; 理解为在圆内任取一点作为弦的中点而作弦，则所求概率为 1/4. 三种不同的理解对应着不同的随机试验，从而有不同的样本点与样本空间，那么所得的答案不同就不足为奇了.

因此，贝特朗奇论不仅没有推翻已有的概率论概念与方法，而且推动了这一理论向前发展. 正是基于这些深刻的讨论，促使从事概率论研究的人们进一步认识到，必须建立概率论的公理化体系，使概率论的发展有一个严格的数学基础. 于是，概率论迎来了本世纪的蓬勃发展时期.

1.3.3 习题

1. 在圆周上任取两点 A, B 连成一弦，再任取两点 C, D 连成一弦. 求 AB 与 CD 相交的概率.

2. 向面积为 S 的 $\triangle ABC$ 内任投一点 P, 求 $\triangle PBC$ 的面积小于 $S/2$ 的概率.

3. 在一张打上正方形格子的纸上投放一枚直径为 1 的硬币. 求方格的边长 a 要多小才能使硬币与方格不相交的概率小于 1%?

4. 甲、乙两船欲停靠同一码头. 假设它们都有可能在某天的一昼夜内任何时刻到达，且甲船与乙船到达后各需在码头停留 3 与 4 小时. 求有船到达时需等待空出码头的概率.

5. 在区间 (0, 1) 内任取两个数，求下列事件的概率: (1) 两数之和小于 1.2; (2) 两数之差的绝对值大于 0.2; (3) 以上二要求全满足.

6. 在区间 $(-1,1)$ 上任取两数 ξ,η, 求二次方程 $x^2+\xi x+\eta=0$ 的二根 (1) 都是实数; (2) 都是正数的概率.

7. 把长为 l 的线段任意折成 3 段, 求 (1) 它们可构成一个三角形; (2) 它们中最长的不超过 $2l/3$ 的概率.

1.4 概率空间

1.4.1 问题的提出

迄今, 我们已经介绍了古典概型与几何概型, 它们分别是具有有限多个样本点与具有不可列无限多个样本点情形的等可能概型. 这是概率论历史上最先接触到的两种概型, 在概率论的建立与发展的初期曾起着重要的作用. 但是它们毕竟局限在 "等可能性" 条件下, 对于具有可列无限多样本点情形及其它许多不具备等可能性的随机试验则不适用. 面对复杂多样的随机现象, 人们发现, 仍象前面两种概型那样逐一具体地定义事件与概率等基本概念是十分困难的, 而必须寻求一个抽象的普遍定义.

另外, 在 1.1.4 节中还曾给出概率的 "统计定义". 尽管借助频率的稳定值引入概率概念时没有等可能性的限制, 但那至多只是一种描述, 而不能作为严格的数学定义. 当时我们就已指出, 不管随机试验能否重复进行, 其中各事件发生的概率总是客观存在的. 大量重复试验中得到的频率的稳定值, 不过是人们认识概率的手段之一. 从理论上讲, 概率的定义应当先于频率的稳定性, 只有当概率的基本理论相当完善之后, 才能彻底讲清 "频率有稳定性" 的确切含义 (参见第四章极限定理). 因此频率的稳定值显然不能作为概率的一般定义.

历史发展到 19 世纪末与 20 世纪初, 一方面由于概率论本身的发展以及来自各应用领域对概率论的需求越来越迫切, 人们感到很有必要完善概率论的基本概念, 使它的发展有一个坚实的数学基础. 另一方面, 当时的数学界有一股公理化的潮流, 许多数学分支都建立起自己的公理化体系, 从而得到迅速的发展. 再加上当时一些相关的数学学科, 如勒贝格测度与积分理论已相当完善. 在这样背景下, 概率

论学者为建立自己的公理化体系所作的努力取得显著的成果. 以著名数学家柯尔莫戈洛夫在 1933 年出版的名著《概率论的基本概念》为主要标志, 概率论终于有了自己的一整套公理体系.

简单地说, 公理化概率论就是测度论式的概率论, 其基本概念间的对照已在 1.1.3 节中由表 1.2 扼要给出. 我们把试验的每一可能结果归结为抽象空间中的点, 样本点所组成的集合就是此试验的随机事件, 而事件发生的概率, 不过是度量这些集合大小的一种特定的测度. 现代概率论教材都是遵照这样的公理体系叙述的. 以下简要介绍概率空间的基本概念与性质, 并给出利用这些性质算概率的进一步的例题.

1.4.2 事件 σ 代数

有关样本空间中哪些子集算作随机事件的问题, 是公理化定义中的第一个议题. 当然, 在不同实际问题中事件的选择可以不尽相同. 例如, 预报某地某日的最低气温, 其样本空间应当是一个区间. 通常人们会关心象 {最低气温为 $0°C$}), {最低气温在 $1°C$ 至 $3°C$ 之间} 这样的结果. 就是说, 诸如单点集与区间这样的子集合应当是随机事件. 然后再经过集合的交、并、余等运算所得集合仍应当是事件. 新产生的事件再经过上述运算又会产生一些事件. 循此下去, 由事件组成的类会相当复杂. 但是, 如果人们只关心会不会结冰, 即只关心 $A=${最低气温在 $0°C$ 以下} 与 $\overline{A}=${ 最低气温高于 $0°C$}, 则由两个事件 A 与 $\overline{A}$ 再经过上述运算, 所生成的事件类却极为简单. 总之, 简单也好, 复杂也好, 只要够用就行. 这当然要根据具体问题而定. 但是, 我们不能逐一讨论各种具体的随机试验, 而是要设法把其中本质的东西抽象出来, 给出随机事件的公理化定义. 人们发现, 作为由样本空间 Ω 的子集组成的事件类, 只要对至多可列的并、交及余运算 (合称**波雷尔运算**) 封闭就够用了.

我们总设样本空间 Ω 是给定的非空集合.

定义 4.1 由样本空间 Ω 的某些子集组成的类 $\mathscr{F}$, 如果具有如

下性质:

(1) $\Omega \in \mathcal{F}$;

(2) 若 $A \in \mathcal{F}$, 则 $\overline{A} \in \mathcal{F}$(对逆封闭);

(3) 若一切 $A_n \in \mathcal{F}$, 则 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$(对可列并封闭).

则称 $\mathcal{F}$ 为 Ω 上的**事件 σ 代数**, 而 $\mathcal{F}$ 中的集合称作**随机事件**, 简称**事件**.

定理 4.1 设 $\mathcal{F}$ 是事件 σ 代数, 则有

(4) $\varnothing \in \mathcal{F}$;

(5) 若 $A_n \in \mathcal{F}, n = 1,2,\cdots$, 则 $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$;

(6) 若 $A_1,\cdots,A_n \in \mathcal{F}$, 则 $\bigcup_{k=1}^{n} A_k \in \mathcal{F}$ 且 $\bigcap_{k=1}^{n} A_k \in \mathcal{F}$;

(7) 若 $A, B \in \mathcal{F}$, 则 $A - B \in \mathcal{F}$.

证明: 由 (1) 与 (2) 立得 (4). 再用对偶律及 (2),(3) 得证 (5). 将 (3) 或 (5) 中 $A_{n+1}, A_{n+2}, \cdots$ 均取作 $\varnothing$ 或 Ω 便得到 (6) 中两个结论. 最后, 由 $A - B = A \bigcap \overline{B}$, 用 (2) 与 (6) 可得 (7). 定理证毕.

至此可知, 由 (1) 至 (3) 定义的事件 σ 代数是对于一切波雷尔运算封闭的非空类, 这对于我们解决各种概率问题已经是足够大了. 注意, 我们的公理体系不考虑不可列无限多个事件的运算.

对于任意给定的样本空间 Ω, 只由 $\varnothing$ 与 Ω 组成的类是事件 σ 代数. 称 $\{\varnothing, \Omega\}$ 为平凡 σ 代数, 它刻画了非随机的必然现象. 另一个极端是由 Ω 的一切子集组成的类, 它对于一切集合的运算均封闭, 因而是事件 σ 代数. 当 Ω 为至多可列集时, 即随机试验只有有限个或可列无限多个结果时, 通常采用这个最大的事件 σ 代数, 即认为 Ω 的任一子集都是事件. 设 A 是样本空间 Ω 的一个非空真子集, 则 $\mathcal{F} = \{\varnothing, A, \overline{A}, \Omega\}$ 满足定义 4.1 中的三个条件, 即此 $\mathcal{F}$ 为事件 σ 代数. 它适用于只有两个可能结果 A(称作成功) 与 $\overline{A}$ (称作失败) 的随机试验.

为给出更复杂的事件 σ 代数的例子, 我们需要引进 "生成 σ 代数" 这个概念. 正如本节开头所说, 可以从 Ω 的一些最基本的子集组

成的类出发，反复经过子集间的波雷尔运算，不断添加新的子集来扩大这个类，直至得到包含那个基本子集类的 σ 代数. 下面换一个方向说，也可以从 Ω 的一切子集组成的最大的 σ 代数出发，清除不必要的子集，逐步缩小这个类，直至得到包含着最基本子集的事件 σ 代数. 设 $\mathcal{G}$ 为 Ω 的子集组成的非空类. 如果 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 都是包含 $\mathcal{G}$ 的 σ 代数, 那么 $\mathcal{F}_1\bigcap\mathcal{F}_2$ 仍然是包含 $\mathcal{G}$ 的 σ 代数. $\mathcal{F}_1\bigcap\mathcal{F}_2\supset\mathcal{G}$ 显然成立. 要证明 $\mathcal{F}_1\bigcap\mathcal{F}_2$ 仍为 σ 代数, 我们逐一验证定义 4.1 中的三条公理即可. 例如由 $\Omega\in\mathcal{F}_1$ 与 $\Omega\in\mathcal{F}_2$ 可知 $\Omega\in\mathcal{F}_1\bigcap\mathcal{F}_2$, 于是公理 (1) 得证. (2) 与 (3) 的验证留作练习. 不难看到，这一论证对于任意多个 σ 代数的交的情形仍然适用. 假定所有包含 $\mathcal{G}$ 的 σ 代数组成一个 σ 代数族 $\{\mathcal{F}_t, t\in T\}$. 由于 Ω 的一切子集组成的类是包含 $\mathcal{G}$ 的 σ 代数, 故上述指标集 T 非空. 按上面的论证, $\bigcap\limits_{t\in T}\mathcal{F}_t$ 仍然是包含 $\mathcal{G}$ 的 σ 代数, 它是族 $\{\mathcal{F}_t, t\in T\}$ 中最小的一个，称之为类 $\mathcal{G}$ **生成的 σ 代数**, 记作 $\sigma(\mathcal{G})$. 有了这些准备、我们可以介绍实数集 $\Re=(-\infty,+\infty)$ 上一个最重要的事件 σ 代数.

例 4.1 波雷尔集 设 $\Omega=\Re$, 取全体半直线组成的类

$$\mathcal{P}=\{(-\infty,x):\ -\infty<x<+\infty\},$$

则称 $\sigma(\mathcal{P})$ **波雷尔集类**, 通常把这个 σ 代数记为 $\mathcal{B}$, 称 $\mathcal{B}$ 中集合为**波雷尔集**.

由 σ 代数的定义及定理 4.1, 注意到

$$[a,b)=(-\infty,b)-(-\infty,a);\quad [a,b]=\bigcap_{n=1}^{\infty}[a,b+\frac{1}{n});$$
$$\{b\}=[a,b]-[a,b);\quad (a,b)=[a,b)-\{a\},$$

可知 $\Re$ 中的一切区间，一切至多可列集，一切开集，一切闭集等都是波雷尔集. 总之，当样本空间 $\Omega=\Re$ 时，取波雷尔集类 $\mathcal{B}$ 作为其上的事件 σ 代数已经够用了.

当 Ω 是 $\Re$ 的某个子区间时，我们可用 Ω 交上述波雷尔子集类 $\mathcal{B}$ 中每一个集，并将得到的类作为 Ω 上的事件 σ 代数.

上述波雷尔集的概念还可以推广到多维情形. 比如平面 $\Re^2$ 上的波雷尔子集类 B^2 可定义为全体矩形所生成的 σ 代数, 它也包含了平面上的常用集合. 在此不一一详述.

1.4.3 概率测度

有了随机事件的概念之后, 我们来建立事件概率的公理化定义. 回忆在概率的古典定义, 几何定义及"统计定义"中, 都强调了它们具有非负性, 规范性和可列可加性. 实际上, 这三条就是事件概率的本质特征, 是我们定义概率的基本公理.

定义 4.2 设 $\mathcal{F}$ 为样本空间 Ω 上的事件 σ 代数, $\mathcal{F}$ 上的实值函数 $P(A)$ 如果具有

1° **非负性**: 任 $A \in \mathcal{F}$ 有 $P(A) \geq 0$;

2° **规范性**: $P(\Omega) = 1$;

3° **可列可加性**: 对 $\mathcal{F}$ 中任何两两不相容的事件列 $\{A_n\}$ 有

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n), \tag{4.1}$$

则称此 P 为 $\mathcal{F}$ 上的**概率测度**, 而 $P(A)$ 称为**事件 A 的概率**.

定义 4.3 一个非空的样本空间 Ω, 一个满足定义 4.1 中条件 (1),(2),(3) 的 Ω 上的事件 σ 代数 $\mathcal{F}$, 一个定义于 $\mathcal{F}$ 上满足定义 4.2 中条件 $1^\circ,2^\circ,3^\circ$ 的概率测度 P, 组成的三元体 $(\Omega,\mathcal{F},P)$ 称为**概率空间**.

每一个随机试验都有它自己的概率空间, 它是研究这个随机试验的概率模型. 概率论的公理化体系中包含着概率空间中共同的规律性的东西, 它们适用于刻画各种具体问题的概率空间.

例 4.2 贝努里概率空间 取 $\mathcal{F} = \{\varnothing, A, \overline{A}, \Omega\}$, 其中 A 为 Ω 的非空真子集. 任取两个正数 p 与 $q, p+q=1$. 令 $P(\varnothing)=0, P(A)=p, P(\overline{A})=q, P(\Omega)=1$. 易证此 P 是一个概率测度. 从而 $(\Omega,\mathcal{F},P)$ 是一个概率空间. 它是描述贝努里试验 (参看 2.2.1 节) 的概率空间.

例 4.3 有限概率空间 样本空间是有限集 $\Omega = \{\omega_1, \cdots, \omega_n\}$. 事件 σ 代数$\mathcal{F}$ 取为 Ω 的一切子集 (共 2^n 个) 组成的类. 取 n 个非负实数 $p_1, \cdots, p_n$, 使 $p_1 + \cdots + p_n = 1$. 最后, 对 Ω 的每一子集 A, 令

$$P(A) = \sum_{\omega_i \in A} p_i, \tag{4.2}$$

不难验证, $\mathcal{F}$ 上的 P 有非负性, 规范性与可列可加性, 即 P 为概率测度. 称此 $(\Omega, \mathcal{F}, P)$ 为一有限概率空间. 特别取 $p_i = 1/n, i = 1, \cdots, n$, 就是古典概率空间. 注意在 1.2.1 节古典概率定义中只提到可加性. 这是因为此时 $\mathcal{F}$ 只含有有限个事件. 故要取出 $3°$ 中两两不相容的事件列 $\{A_n\}$, 其中必有无限多个是不可能事件 $\varnothing$. 由下面的定理 4.2 便得, 此时可加性与可列可加性是等价的.

例 4.4 离散概率空间 样本空间 $\Omega = \{\omega_1, \omega_2, \cdots\}$,$\mathcal{F}$ 仍取为 Ω 一切子集所组成的类. 取非负实数列 $\{p_n\}$, 使 $\sum\limits_{n=1}^{\infty} p_n = 1$. 再按 (4.2) 式定义 $P(A)$. 可以验证此 P 具有性质 $1°, 2°, 3°$. 从而 $(\Omega, \mathcal{F}, P)$ 为概率空间, 称为离散概率空间.

例 4.5 几何概率空间(以 1 维为例) 样本空间 Ω 是 $(-\infty, +\infty)$ 中的波雷尔集, 具有正的有限勒贝格测度. 事件 σ 代数$\mathcal{F}$ 取作 Ω 中的波雷尔集类. 对每个 $A \in \mathcal{F}$, 取 $P(A)$ 如上节 (3.1) 式. 不难验证 P 具有性质 $1°, 2°, 3°$. 于是得到几何概型的概率空间 $(\Omega, \mathcal{F}, P)$.

下面从概率空间的公理 $1°, 2°, 3°$ 出发, 导出一般概率空间的普遍性质.

定理 4.2 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间, 则其上概率 P 具有如下性质:

$4°$ $P(\varnothing) = 0$;

$5°$ 有限可加性: 若 $A_1, \cdots, A_n$ 为 $\mathcal{F}$ 中两两不相容事件, 则

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k); \tag{4.3}$$

6° 可减性：若 $A, B \in \mathcal{F}$ 且 $A \subset B$，则 $P(B-A) = P(B)-P(A)$；

7° 单调性：若 $A, B \in \mathcal{F}$ 且 $A \subset B$，则 $P(A) \leq P(B)$；

8° $P(\overline{A}) = 1 - P(A)$；

9° 加法定理：任 $A_1, \cdots, A_n \in \mathcal{F}$ 有

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) - \sum_{1\leq i<j\leq n} P(A_iA_j)$$

$$+ \sum_{1\leq i<j<k\leq n} P(A_iA_jA_k) - \cdots + (-1)^{n-1}P(A_1\cdots A_n); \tag{4.4}$$

10° 下连续性：若 $A_n \in \mathcal{F}$ 且 $A_n \subset A_{n+1}, n = 1, 2, \cdots$，则

$$P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n} P(A_n);$$

11° 上连续性：若 $A_n \in \mathcal{F}$ 且 $A_n \supset A_{n+1}, n = 1, 2, \cdots$，则

$$P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n} P(A_n);$$

12° 次可加性：任 $A_n \in \mathcal{F}, n = 1, 2, \cdots$，有

$$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

证明：3° 中取 $A_n \equiv \emptyset$ 即得 4°. 3° 中取 $A_{n+1} = A_{n+2} = \cdots = \emptyset$，联合 4° 便得 5°. 注意到 $A \subset B$ 时有 $B = A\bigcup(B-A)$，且右方为不交并，用 5° 便得证 6°. 由 6° 及 1° 得证 7°. 5° 中取 $B = \Omega$ 再用 2° 可得 8°.

为证加法定理 9°，我们对 n 作归纳法. $n = 2$ 时将 $A_1\bigcup A_2$ 写为不交并

$$A_1\bigcup A_2 = [A_1 - (A_1A_2)]\bigcup(A_1A_2)\bigcup[A_2 - (A_1A_2)].$$

运用 5°, 6° 便得证 (4.4) 式在 $n = 2$ 时真. 现设 (4.4) 在 n 时已真，将 $\bigcup\limits_{k=1}^{n+1} A_k$ 视为 $\bigcup\limits_{k=1}^{n} A_k$ 与 A_{n+1} 两个事件之并. 运用 $n = 2$ 时 (4.4)

已真及归纳法假设，直接计算便可推出 (4.4) 式于 $(n+1)$ 时仍真. 性质 9° 得证.

在 $\{A_n\}$ 单调上升的情形，令 $B_1=A_1, B_n=A_n-A_{n-1}(n>1)$. 则有 $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$, 且右方为不交并，运用 3° 得

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n} \sum_{k=1}^{n} P(B_k).$$

再运用 5° 及 $A_n = \bigcup_{k=1}^{n} B_k$ 便得证

$$P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n} P(\bigcup_{k=1}^{n} B_k) = \lim_{n} P(A_n).$$

此即下连续性 10°. 在 $\{A_n\}$ 为下降事件列的情形考虑事件列 $\{\overline{A}_n\}$, 则由 10° 及 8° 可导出上连续性 11°. 最后证明次可加性 12°. 利用加法定理 9° 并对 n 作归纳法可得，对任何 $n \geq 1$ 有

$$P(\bigcup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} P(A_k).$$

令 $n \to \infty$, 据概率的下连续性 10° 便完成 12° 的证明. 定理至此全部证完.

我们指出，上述性质 10° 与 11° 所以称为"连续性"，是因为它们都表明在单调情形下，概率与极限符号可以交换次序. 例如对上升事件列情形，我们有 (参见习题 1.1.5 的第 5 题)

$$\overline{\lim_{n}}\, A_n = \underline{\lim_{n}}\, A_n = \bigcup_{n=1}^{\infty} A_n.$$

这表明事件列 $\{A_n\}$ 有极限 $\lim_{n} A_n = \bigcup_{n=1}^{\infty} A_n$. 于是下连续性 10° 可改写为 $P(\lim_{n} A_n) = \lim_{n} P(A_n)$.

1.4.4 利用性质算概率的例

前面我们已经给出直接利用古典概率及几何概率的定义计算概率的例题. 现在我们已对一般的概率空间导出了概率的一系列性质,

这些性质当然也适用前述两种特殊的概率空间. 下面介绍利用概率的性质 4° 至 12° 计算事件概率的例子.

例 4.6 n 对夫妇任意在一排 $2n$ 个椅子上就座, 求事件 $A = \{$有夫妇不相邻$\}$ 的概率.

解: 显然样本点是这 $2n$ 个人的全排列, 即 $n(\Omega) = (2n)!$. "有夫妇不相邻" 包含了有一对, 两对, 等情况, 有一对情形又要分是这一对还是那一对, 以及不相邻夫妇的间隔多远等等. 故直接考虑事件 A 中的样本点数较为困难. 而其对立事件 $\overline{A} = \{$ 无夫妇不相邻 $\}$ 则很简单. 由乘法原理易得 $n(\overline{A}) = n!2^n$. 故 $P(\overline{A}) = 1/(2n-1)!!$. 用概率性质 8° 立得 $P(A) = 1 - 1/(2n-1)!!$.

运用性质 8° 算概率是一个有效的方法. 当我们要计算任一个事件 A 的概率时, 都可以先想一下, 是 A 还是 $\overline{A}$ 的概率容易求出? 往往可以得到很好的效果.

例 4.7 蒲丰投针问题的续 向画满间隔为 a 的平行直线的桌面上任投一直径为 $l(l < a)$ 的半圆形纸片, 求事件 $E = \{$ 纸片与某直线相交 $\}$ 的概率.

解: 可以象解例 3.3 那样直接建立这个问题的几何概率模型. 下面设法用概率的加法定理 (性质 9°) 得到所求的概率.

设想将半圆形纸片拼成一个圆形, 并记 $F = \{$ 新拼的半圆形与某直线相交 $\}$. 我们有

$E \bigcup F = \{$ 直径为 l 的圆形与某直线相交 $\}$,

$E \bigcap F = \{$ 长为 l 的线段 (即公共直径) 与某直线相交 $\}$.

由例 3.3 知

$$P(E \bigcap F) = \frac{2l}{\pi a}.$$

为求 $P(E \bigcup F)$, 以 r 表圆心到桌面上最近直线的距离. 则 $\Omega = [0, a/2]$ 而 $E \bigcup F = \{r \in \Omega : r \leq l/2\}$. 故

$$P(E \bigcup F) = \frac{l}{a} = \frac{\pi l}{\pi a}.$$

利用加法定理 (4.4) 式得

$$P(E)+P(F)=P(E\bigcup F)+P(E\bigcap F)=\frac{\pi l+2l}{\pi a}.$$

再由对称性知 $P(E)=P(F)$, 故所求概率

$$P(E)=\frac{\frac{1}{2}\pi l+l}{\pi a}.$$

注意上式的分子恰为半圆形纸片的周长. 其实在最初算出的针与某直线相交的概率中, 分子 $2l$ 也可看作针的周长. 同时, 上面算出圆与某直线相交的概率为 $\frac{\pi l}{\pi a}$, 其分子也是圆周长. 我们不禁要问, 抛掷更一般的图形时是否也有相应的结论呢? 回答是肯定的. 在此不详加叙述, 只在本节习题 8 中讨论另外一个特例.

例 4.8　配对问题 参加集合的 n 个人将他们的帽子放在一起, 会后每人任取一顶帽子戴上. 求恰有 k 个人戴对自己的帽子的概率.

解: 为叙述简便, 我们把 "1 个人戴对了自己的帽子" 简称为 "1 个配对", 并记 $A_k=\{$恰有 k 个配对$\}$.

先看 $k=0$ 的情形, 即求 $A_0=\{n$ 个人中无配对$\}$ 的概率. 初看起来 A_0 比其对立事件 $\overline{A}_0=\{n$ 个人中有配对$\}$ 要简单. 正象我们前面说过的, "有配对" 要分为有 1 个, 2 个配对等情况, 有 1 个配对又要分析是哪个人配对等等. 但是事件 $\overline{A}_0$ 可以分解为若干事件之并, 而加法定理 9° 提供了分解复杂 (较难求概率) 事件为简单 (易求概率) 事件来计算概率的途径. 令 $B_i=\{$ 第 i 个人配对 $\}, i=1,\cdots,n$. 则有 $\overline{A}_0=\bigcup\limits_{i=1}^{n}B_i$. 显然右方诸 B_i 不是两两不相容的, 故不能用有限可加性计算 $P(\overline{A}_0)$. 但是可运用加法定理 9° 得到

$$P(\overline{A}_0)=\sum_{i=1}^{n}P(B_i)-\sum_{1\le i<j\le n}P(B_iB_j)+\cdots+(-1)^{n-1}P(B_1\cdots B_n).$$

不妨设 n 顶帽子已排放完毕, 样本点就是 n 个人的全排列, 即 $n(\Omega)=n!$ 易见, 对一切 $i,j,k=1,2,\cdots,n$, 只要 i,j,k 两两不同, 则有

$$n(B_i)=(n-1)!,\quad n(B_iB_j)=(n-2)!,$$

$$n(B_iB_jB_k) = (n-3)!, \cdots, n(B_1 \cdots B_n) = 0! = 1,$$

代入可得到

$$P(\overline{A}_0) = \sum_{1 \leq i \leq n} \frac{1}{n} - \sum_{1 \leq i < j \leq n} \frac{1}{n(n-1)} + \cdots + (-1)^{n-1} \frac{1}{n!}.$$

注意到上式各和号内被加项均与 i, j 等指标无关，且第 k 个和号恰含 $\binom{n}{k}$ 项. 代入可得

$$P(\overline{A}_0) = \frac{1}{1!} - \frac{1}{2!} + \cdots + (-1)^{n-1} \frac{1}{n!}.$$

再用性质 8° 得

$$P(A_0) = 1 - P(A_0) = \sum_{i=0}^{n} (-1)^i \frac{1}{i!}.$$

注意，当人数 n 趋于无穷时，"无配对"概率的极限为 e^{-1}. 这是客观世界中存在的又一个与 e 有关的量.

下面对一般的 $k \geq 1$ 求 $P(A_k)$. 为此记 $C_k = \{$ 恰好某指定 k 个人配对 $\}$. 由乘法原理可得 $n(A_k) = \binom{n}{k} \cdot n(C_k)$. 注意到恰好某 k 个人配对相当于其余 $n-k$ 个无配对，由上述对 A_0 所得结果知

$$P(C_k) = \sum_{i=0}^{n-k} (-1)^i \frac{1}{i}.$$

注意此时共有 $n-k$ 个人，故上述概率等于 $n(C_k)/(n-k)!$, 由此可得

$$n(C_k) = (n-k)! \sum_{i=0}^{n-k} (-1)^i \frac{1}{i}.$$

我们最终得到

$$P(A_k) = \frac{\binom{n}{k} n(C_k)}{n!} = \frac{1}{k!} \sum_{i=0}^{n-k} (-1)^i \frac{1}{i}.$$

此结果对于 $k=0,1,\cdots,n$ 全成立. 令人数 $n\to\infty$ 得极限概率为 $e^{-1}/k!$.

本例解题过程中用到: (1) 通过对立事件 $\overline{A}_0$ 算概率 $P(A_0)$; (2) 分解事件 $\overline{A}_0$ 为 $B_1,\cdots,B_n$ 之并, 用加法定理化难为易计算概率; (3) 通过已有的概率 $P(C_k)$ 得到事件 C_k 中样本点个数. 这些都是很有启发性的方法.

下面的例题提供了概率的上连续性 11° 与次可加性 12° 的应用.

例 4.9 假设有一个无限大的箱子和标号为自然数的无限多个球, 并且有一个装置可以"要多快有多快"地完成放球取球动作. 假定我们在差 1 分 12 点时, 将 1 至 10 号球放入箱中并取出 1 号球, 在差 1/2 分 12 点时将 11 至 20 球放入箱中并取出 11 号球, $\cdots$, 如此下去, 那么在 12 点整, 箱中将有无限多个球. 现保持上述放球方式不变, 而改变取球的号码为: 第 1 次取出 1 号球, 第 2 次 (即差 1/2 分 12 点时) 取出 2 号球, $\cdots$, 如此下去, 则在 12 点整箱子应是空的. 如果再把取球方式改为"每次任取出一球", 完整地说, 对 $n=1,2,\cdots$, 在差 $1/2^{n-1}$ 分 12 点时将第 $(10n-9)$ 号至第 $10n$ 号球放入箱中, 并任取出一球. 我们要证明, 事件 $A=\{$ 在 12 点整箱子是空的 $\}$ 发生的概率为 1.

证明: 令 $F_i=\{$ 在 12 点整 i 号球在箱中 $\}$, 则 $\overline{A}=\bigcup\limits_{i=1}^{\infty}F_i$. 用次可加性知

$$P(\overline{A})\leq\sum_{i=1}^{\infty}P(F_i).$$

如能对 $i=1,2,\cdots$, 证明 $P(F_i)=0$, 则由上式及非负性 1° 可知 $P(\overline{A})=0$, 从而完成了证明. 我们先以 F_1 为例.

以 E_n 表示在前 n 次放球取球后 (即差 $1/2^{n-1}$ 分 12 点时)1 号球仍在箱中这一事件. 则有

$$P(E_n)=\frac{9\cdot18\cdot27\cdot\cdots\cdot(9n)}{10\cdot19\cdot28\cdot\cdots\cdot(9n+1)}.$$

这是因为，为要 E_n 发生必须且只需前 n 次都没取到 1 号球，于是第 1 次有 9 种取法，第 2 次有 18 种取法，$\cdots$，第 n 次有 $9n$ 次取法. 此即 E_n 中样本点数. 同理可得样本点总数为 $10\cdot 19\cdot 28\cdot(9n+1)$. 注意到事件列 $\{E_n\}$ 是单调下降的，有

$$\lim_n E_n = \bigcap_{n=1}^{\infty} E_n = F_1,$$

故用概率的上连续性 11° 可得

$$P(F_1) = \lim_n P(E_n) = \prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right).$$

由分析知级数 $\sum\limits_{n=1}^{\infty}\frac{1}{9n} = +\infty$，这等价于 $\prod\limits_{n=1}^{\infty}(1+\frac{1}{9n}) = -\infty$，故

$$P(F_1) = \left[\prod_{n=1}^{\infty}(1+\frac{1}{9n})\right]^{-1} = 0.$$

对于一般的 i，只要 n 充分大，使得 i 号球已放入箱中，其余计算类似，即得证 $P(F_i) = 0$. 至此完成了 $P(A) = 1$ 的证明.

1.4.5 习题

1. 袋中有编号为 1 至 6 的 6 个球，现不放回地任取出 3 个，求 3 个球中号码 1,2,3 至少出现 1 个的概率.

2. 袋中有 $(N-1)$ 个白球与 1 个黑球，每次从袋中任取出 1 球，并换一个白球放回. 求第 k 次取出白球的概率.

3. n 对夫妇任意围一圆桌就座，求有夫妇不相邻的概率.

4. 将一枚均匀的骰子掷 n 次，求得到的最大点数为 5 的概率.

5. 从装有红、白、黑球各一个的袋中任意有放回地取球，直至 3 种颜色的球都取出过为止. 求取球次数 (1) 大于 k; (2) 恰为 k 的概率.

6. 从一副 52 张扑克牌中有放回地任取 n 张 $(n\geq 4)$，求这 n 张包含了全部 4 种花色的概率.

7. 从 $0, 1, \cdots, 9$ 这 10 个数字中不放回地任取 n 个，求 n 个数之积能被 10 整除的概率.

8. 向画满间隔为 a 的平行直线的桌面上任投放一三角形. 假定三角形三条边长 l_1, l_2, l_3 均小于 a. 求此三角形与某直线相交的概率.

9. 一次口试准备了 N 张考签. 每个考生从中任取一张用毕放回，求 n 个 $(n \geq N)$ 考生考完后，考签全被用过的概率.

10. 用概率想法求 n 阶行列式展开式中含主对角线元的项数.

11. 用概率想法证明对任何自然数 $a, A (a < A)$ 有

$$\begin{aligned}\frac{A}{a} &= 1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} \\ &\quad + \cdots + \frac{(A-a)(A-a-1)\cdots 2 \cdot 1}{(A-1)(A-2)\cdots(a+1)a}.\end{aligned}$$

12. 试证明，对任意两个事件 A 与 B 有

$$P(A) \vee P(B) \leq P(A \bigcup B) \leq 2[P(A) \vee P(B)].$$

13. 设 $P(A) = P(B) = 1/2$. 求证 $P(AB) = P(\overline{A}\,\overline{B})$.

14. 求证对任意 n 个事件 $A_1, \cdots, A_n$ 有

$$P(\bigcap_{k=1}^{n} A_k) \geq \sum_{k=1}^{n} P(A_k) - n + 1.$$

15. 甲乙轮流掷一均匀硬币，甲先掷，直到某人掷出一正面为止，掷出正面者获胜. (1) 试写出此随机试验的概率空间；(2) 求甲获胜的概率；(3) 求至多掷 n 次的概率.

16. 试证：

$$P(\varliminf_{n} A_n) \leq \varliminf_{n} P(A_n), \quad \varlimsup_{n} P(A_n) \leq P(\varlimsup_{n} A_n).$$

17. 设 $\mathcal{O}$ 为 $\Re$ 上的开集全体，$\mathcal{J}$ 为其上有理顶点开区间全体. 求证 $\Re$ 中的波雷尔集类 $\mathcal{B} = \sigma(\mathcal{O}) = \sigma(\mathcal{J})$.

1.5 条件概率

1.5.1 定义与乘法定理

在谈及随机试验及其中各个事件的概率的时候, 总是在一组确定的条件下讨论的. 本书开篇所介绍的随机现象, 就是"在一定条件下进行试验或观测, 其结果不能完全预言"的现象. 这里的条件是整个试验的共同条件, 是我们讨论问题的大前提. 有时, 人们还需要讨论在某些附加条件下的试验结果. 这些附加条件 (即小前提) 通常以"某个事件已经发生"的形式给出. 这就是已知某事件已发生后的条件概率.

先看一个古典概率的例子. 某人从 1 至 10 这 10 个数中任取一数 (大前提), 那么事件 $A = \{$ 取到的数比 3 大 $\}$ 中包含着上述 10 个样本点中的 7 个. 从而 $P(A) = 7/10$. 现在, 如果已经得知取到的数是偶数 (小前提), 即事件 $B = \{$ 取到偶数 $\}$ 已经发生, 则因为尚在考虑之中的事件 B 中只含 5 个等可能样本点, 使 A 出现的有其中的 4 个, 故 A 的概率上升为 $4/5$. 此即已知 B 发生后 A 的条件概率, 记为 $P(A|B) = 4/5$. 注意到 $n(B) = 5$, $n(AB) = 4$. 故有

$$P(A|B) = \frac{n(AB)}{n(B)} = \frac{n(AB)/n(\Omega)}{n(B)/n(\Omega)} = \frac{P(AB)}{P(B)},$$

这启发我们以 $P(AB)$ 与 $P(B)$ 之比作为条件概率 $P(A|B)$ 的一般定义.

定理 5.1 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间, $B \in \mathcal{F}$ 满足 $P(B) > 0$. 对任何事件 $A \in \mathcal{F}$, 取

$$P(A|B) = \frac{P(AB)}{P(B)}, \tag{5.1}$$

则作为 $\mathcal{F}$ 上之 A 的集函数, $P(A|B)$ 仍满足概率定义中的 $1°$ 非负性, $2°$ 规范性和 $3°$ 可列可加性. 称此 $P(A|B)$ 为已知事件 B 发生后 A 的**条件概率**.

证明: 非负性与规范性显然成立. 对于 $\mathcal{F}$ 中任一两两不相容事件列 $\{A_n\}$, 由于序列 $\{A_nB\}$ 仍两两不相容, 故由 P 的可列可加性可得

$$P(\bigcup_{n=1}^{\infty}[A_nB]) = \sum_{n=1}^{\infty} P(A_nB).$$

两边同时除以 $P(B) > 0$ 便得证可列可加性:

$$P(\bigcup_{n=1}^{\infty} A_n|B) = \sum_{n=1}^{\infty} P(A_n|B).$$

此定理表明 (5.1) 式给出的集函数 $P(A|B), A \in \mathcal{F}$ 仍是 $(\Omega, \mathcal{F})$ 上的一个概率. 因此, 我们已经与即将导出的概率测度的一切性质同样适用于条件概率. 此外我们指出, 在条件概率的定义中要求作为条件的事件有正概率, 形式上是为使 (5.1) 式中分母不为 0. 在实践中人们也不以零概率事件 (例如不可能事件) 的发生作为讨论问题的前提.

将 (5.1) 式变形可得

$$P(AB) = P(B)P(A|B), \tag{5.2}$$

此式给出了求两个事件同时发生之概率的公式. 它可以推广到多个事件的情形.

定理 5.2　乘法定理 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间, $A_k \in \mathcal{F}, k = 1, 2, \cdots, n$. 如果 $P(A_1 \cdots A_{n-1}) > 0$, 则有

$$\begin{aligned} P(A_1A_2\cdots A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots \\ &\quad P(A_n|A_1\cdots A_{n-1}). \end{aligned} \tag{5.3}$$

证明: 条件 $P(A_1 \cdots A_{n-1}) > 0$ 保证 (5.3) 式右方诸条件概率有意义. 再将 (5.1) 式代入这些条件概率即得证 (5.3) 式.

作为乘法定理的一个应用, 我们重解结绳问题.

例 5.1　将 n 根绳的 $2n$ 个头任意两两相接, 求事件 $A = \{$ 恰结成 n 个圈 $\}$ 的概率.

解: 以 B_i 表示第 i 根绳的头与尾恰好相接, $i=1,\cdots,n$. 则有 $A=B_1B_2\cdots B_n$. 我们约定将这 $2n$ 个头任排为一列, 然后第 1 与第 2 个头相接, 第 3 个头与第 4 个头相接, $\cdots$, 第 $2n-1$ 个头与第 $2n$ 个头相接. 于是一个样本点相当于 $2n$ 个头的一个全排列, $n(\Omega)=(2n)!$. 而 $B_1=\{$ 第 1 根绳的首尾相接 $\}$ 中的样本点数为 $n(B_1)=2n(2n-2)!$, 故 $P(B_1)=1/(2n-1)$. 下考虑 $P(B_2|B_1)$, 因已知第 1 根绳已头尾相接, 可以将它舍弃. 于是 $P(B_2|B_1)$ 化为 $(n-1)$ 根绳时第 2 根绳的头尾相接的概率, 故 $P(B_2|B_1)=1/[2(n-1)-1]=1/(2n-3)$. 同理 $P(B_3|B_1B_2)=1/(2n-5)$, 循此下去, 运用乘法定理 (5.3) 式便可得到

$$\begin{aligned}P(A) &= P(B_1B_2\cdots B_n)\\ &= P(B_1)P(B_2|B_1)\cdots P(B_n|B_1\cdots B_{n-1})\\ &= \frac{1}{(2n-1)!!}.\end{aligned}$$

在下面的 1.5.3 节我们将给出结绳问题的第三种解法.

1.5.2 全概公式与贝叶斯公式

将复杂问题适当地分解为若干个简单问题而逐一解决, 是人们常用的工作方法. 在解决较复杂的求概率问题时, 人们也希望把所涉及的复杂事件分解为简单事件之并. 例如在二维几何概型中, 求各事件发生的概率被归结为求平面图形的面积. 而我们早就知道, 在求面积时可以通过 "割补法" 将较复杂的图形 (例如多边形) 分解为若干个简单的图形 (三角形). 本节介绍的全概公式, 就是借助引入各种小前提, 将样本空间适当地分解为若干部分, 使得在每个部分中 (即在各小前提下), 容易求得所要的概率. 先引进样本空间的分割这一概念.

定义 5.1 设 $(\Omega,\mathcal{F},P)$ 为一概率空间, $\{B_n\}$ 为 $\mathcal{F}$ 中两两不相容事件列 (有限或可列无限多个), 使得对一切 n 有 $P(B_n)>0$, 且 $\bigcup\limits_n B_n=\Omega$, 则称此 $\{B_n\}$ 为 Ω 的一个**分割**.

定理 5.3 全概公式 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，而 $\{B_n\}$ 为 Ω 的一个分割，则对一切 $A \in \mathcal{F}$，有

$$P(A) = \sum_n P(B_n)P(A|B_n). \tag{5.4}$$

证明：由分割的定义可知 $A = \bigcup_n (AB_n)$，且右方为两两不相容事件之并. 利用可列可加性及乘法定理 (5.2) 式便得证

$$P(A) = \sum_n P(AB_n) = \sum_n P(B_n)P(A|B_n).$$

我们强调指出，分割 $\{B_n\}$ 是为了计算 $P(A)$ 而人为地引入的. 选择适当可以使计算大为简化. 选择不适当，则不利于问题的解决. 这一点是初学者应当格外留心的.

例 5.2 某工厂的 1,2,3 车间生产同一种产品，产量依次占 1/2, 1/4, 1/4，而次品率分别为 0.01, 0.01 及 0.02. 现从这个厂的产品中任取出一件，求 $A = \{$ 取到 1 件次品 $\}$ 的概率.

解：问题在于不知道取到的产品是哪个车间生产的. 如果令 $B_i = \{$ 取到 i 车间的产品 $\}$，$i = 1, 2, 3$，则 B_1, B_2, B_3 为 Ω 的一个分割. 由题意知 $P(B_1) = 0.5, P(B_2) = P(B_3) = 0.25, P(A|B_1) = 0.01, P(A|B_2) = 0.01, P(A|B_3) = 0.02$. 用全概公式 (5.4) 式立得

$$\begin{aligned} P(A) &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ &= 0.0125. \end{aligned}$$

例 5.3 一般摸球模型 袋中有 r 个红球与 b 个黑球. 每次从袋中任摸出 1 球并连同 s 个同色球一起放回. 以 R_n 表示第 n 次摸出红球，试证 $P(R_n) = r/(r+b)$.

证明：我们对摸球次数 n 作归纳法. $n = 1$ 时 $P(R_1) = r/(r+b)$ 显然成立. 假设 $(n-1)$ 时成立. 为求 $P(R_n)$，我们以第 1 次取球的

可能结果 R_1 与 $\overline{R}_1 = \{$ 第 1 次取出黑球 $\}$ 作为 Ω 的分割, 用全概公式可得

$$P(R_n) = P(R_1)P(R_n|R_1) + P(\overline{R}_1)P(R_n|\overline{R}_1).$$

注意在 R_1 条件下, 袋中有 $r+s$ 个红球与 b 个黑球. 而 $P(R_n|R_1)$ 相当于自 $r+s$ 个红球与 b 个黑球出发, 在第 $n-1$ 次摸出红球的概率, 由归纳法假设有 $P(R_n|R_1) = (r+s)/(r+s+b)$. 同理, $P(R_n|\overline{R}_1) = r/(r+b+s)$. 代入可得

$$P(R_n) = \frac{r}{r+b} \cdot \frac{r+s}{r+s+b} + \frac{b}{r+b} \cdot \frac{r}{r+b+s} = \frac{r}{r+b}.$$

这是最一般的摸球模型. 当 $s=0$ 时相当于放回摸球, 而 $s=-1$ 相当于不放回摸球, 于是, 又一次证明了抽签是公平的 (见例 2.5).

为了引进贝叶斯公式, 我们继续讨论例 5.2. 假定事件 $A = \{$ 取到 1 件次品 $\}$ 已经发生. 这就产生了分担由此而引起的责任问题. 显然, 各车间应承担的责任应当取决于已被查出的次品是他们车间生产的概率 $P(B_i|A), i = 1,2,3$. 这个条件概率不难由条件概率的定义及乘法定理得出. 例如,

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{0.5 \times 0.01}{0.0125} = 0.4.$$

定理 5.4　贝叶斯公式 设 $(\Omega, \mathcal{F}, P)$ 为一个概率空间, $\{B_n\}$ 为 Ω 的一个分割. 则对 $\mathcal{F}$ 中任何事件 A, 只要 $P(A) > 0$ 就有

$$P(B_n|A) = \frac{P(B_n)P(A|B_n)}{\sum\limits_i P(B_i)P(A|B_i)}. \tag{5.5}$$

证明: 用 (5.1), (5.2) 及 (5.4) 式立得.

我们介绍贝叶斯公式的一个简单应用.

例 5.4 假设通过验血诊断某种疾病的误诊率仅为 5%. 就是说, 如记 $A = \{$验血结果为阳性$\}$, $B = \{$ 受检者患此病 $\}$, 则 $P(\overline{A}|B) = P(A|\overline{B}) = 5\%$. 如果受检人群中仅有 0.5% 患有此病, 即 $P(B) = 0.005$, 求一个化验为阳性的人确患此病的条件概率 $P(B|A)$.

解: 以 B 与 $\overline{B}$ 为分割运用贝叶斯公式可得

$$
\begin{aligned}
P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(\overline{B}|P(A|\overline{B})} \\
&= \frac{0.005\times 0.95}{0.005\times 0.95+0.995\times 0.05}=0.087.
\end{aligned}
$$

这个结果出人意料地小. 因为一般人们会以为, 既然这种验血技术这么可靠, 那么验血结果为阳性的人, “十有八九” 是患病无疑了. 而我们计算的结果却表明, 化验为阳性的人患此病的概率还不足百分之九. 从以上算式不难看到, 式中分母的两项中, 第一项差不多是第二项的十分之一, 而分子又与分母第一项相同, 因此这个概率差不多是十一分之一. 就是说, 所求概率如此小的根本原因是这种疾病的患病率 $P(B)=0.005$ 很小.

贝叶斯公式是英国哲学家贝叶斯于 1763 年首先提出的. 假定 B_1, $B_2,\cdots$ 是某个过程的若干可能的前提, 则 $P(B_i)$ 是人们事先对各前提条件出现可能性大小的估计, 称之为验前概率. 如果这个过程得到了一个结果 A, 那么贝叶斯公式提供了我们根据 A 的出现而对各前提条件作出新评价的方法. $P(B_i|A)$ 即是对前提 B_i 出现概率的重新认识, 称 $P(B_i|A)$ 为验后概率. 经过多年的发展与完善, 贝叶斯公式以及由此发展起来的一整套理论与方法, 已经形成为概率统计中的一个冠以 “贝叶斯” 的学派, 在自然科学及国民经济的许多领域中有着广泛应用. 下面我们仅以疾病诊断为例, 介绍贝叶斯决策的基本思想. 由病历统计可得到某地区在指定时间内患感冒 (B_1) 、患结核 (B_2) 及患风湿 (B_3) 等疾病的概率, 这就是验前概率 $P(B_i)$. 再者, 根据病理学及病历资料, 可以确定上述疾病的患者出现 “发烧”(A) 这一症状的概率 $P(A|B_i)$. 于是, 人们可以利用贝叶斯公式, 很快地算出诸 B_i 的验后概率 $P(B_i|A)$. 这样, 当医生面对一个有症状 A(发烧) 的病人时, 他就可以根据已经算出的 $P(B_i|A)$, 择其较大者作出判断. 当然, 实际应用过程不会如此简单. 比如, 医生还需要对患者的其它体征加以综合考虑等等. 但是, 在计算机已广泛应用的今天, 这种数学方法的实用价值已大大增强.

1.5.3 条件化及计算概率的递推方法

全概公式是概率论发展前期的一个里程碑. 它可以将较难计算的无条件概率转化为简单易算的条件概率. 特别是在计算与自然数 n 有关的事件概率时, 通过将样本空间进行适当的分割, 使无条件概率条件化, 往往可以得到所求概率的一个递推公式 (或称差分方程), 从而达到较容易地算出概率的目的. 我们将通过例题介绍这一方法.

例 5.5 结绳问题 将 n 根绳的 $2n$ 个头任意两两相接, 求事件 $A_n=\{$ 恰接成 n 个圈 $\}$ 的概率.

解: 记 $P(A_n)=p_n$. 以第 1 根绳的头连接情况作为条件, 即以 $B=\{$ 第 1 根绳的头接在自己的尾上 $\}$ 与 $\overline{B}$ 为分割用全概公式可得

$$p_n=P(B)P(A_n|B)+P(\overline{B})P(A_n|\overline{B}),$$

由于 $P(B)=1/(2n-1)$ (参见前面的例 5.1), 而 $P(A_n|B)=P(A_{n-1})=p_{n-1}$, $P(A_n|\overline{B})=0$. 代入便得递推公式

$$p_n=\frac{1}{2n-1}p_{n-1},\quad n=2,3,\cdots.$$

反复用此公式, 并联合初始值条件 $p_1=1$ 可得

$$p_n=\frac{1}{(2n-1)!!}.$$

这与前面用古典概型与乘法定理所得的结果 (参见例 2.7 与例 5.1) 是一致的.

例 5.6 甲、乙轮流掷一枚均匀的骰子. 甲先掷, 以后每当某人掷出 1 点时则交给对方掷, 否则此人继续掷. 试求事件 $A_n=\{$ 第 n 次由甲掷 $\}$ 的概率.

解: 记 $p_n = P(A_n)$. 以事件 A_{n-1} 及 $\overline{A}_{n-1}$ 为 Ω 的分割用全概公式可得

$$p_n = P(A_n) = P(A_{n-1})P(A_n|A_{n-1}) + P(\overline{A}_{n-1})P(A_n|\overline{A}_{n-1}).$$

现 $P(A_{n-1}) = p_{n-1} = 1 - P(\overline{A}_{n-1})$, $P(A_n|A_{n-1}) = 5/6$, $P(A_n|\overline{A}_{n-1}) = 1/6$, 代入上式可得

$$p_n = \frac{5}{6}p_{n-1} + \frac{1}{6}(1 - p_{n-1}) = \frac{2}{3}p_{n-1} + \frac{1}{6},$$

将它化为便于递推的形式

$$p_n - \frac{1}{2} = \frac{2}{3}\left(p_{n-1} - \frac{1}{2}\right), \ n = 2, 3, \cdots.$$

反复用此递推公式并联合初始条件 $p_1 = 1$ 便得

$$p_n - \frac{1}{2} = \left(\frac{2}{3}\right)^{n-1}\left(p_1 - \frac{1}{2}\right) = \frac{1}{2}\left(\frac{2}{3}\right)^{n-1},$$

从而有

$$p_n = P(A_n) = \frac{1}{2}\left[1 + \left(\frac{2}{3}\right)^{n-1}\right], \ n = 1, 2, \cdots.$$

在上面两个例题中, 所得到的递推公式只涉及 n 与 $n-1$ 这两个相邻的时刻. 下面再介绍一个涉及三个相邻时刻的例子. 同时引进随机徘徊这一重要的概率模型.

例 5.7 随机徘徊的吸收概率 质点在数轴的整点上运动. 无论它处在哪一个点 i 上, 下一时刻都以概率 p 向右移动到 $(i+1)$, 以概率 q 向左移动到 $(i-1)$, $p, q > 0$, $p + q = 1$. 我们把质点的这种运动称为 (直线上的) **随机徘徊**. 现在考虑数轴上两个特殊的点 0 与 a $(a > 1)$, 假定质点运动到 0 或 a 之后就永远不再移动, 称这样的 0 与 a 为随机徘徊的**吸收壁**. 现求自 i $(0 < i < a)$ 出发的质点将被 0 或 a 吸收的概率.

解: 令 $E_i = \{$ 质点自 i 出发 $\}$, $F = \{$ 质点将在 0 被吸收 $\}$, $G = \{$ 质点将在 a 被吸收 $\}$. 下面对 $i = 0, 1, \cdots, a$ 求

$$P_i = P(F|E_i) \quad 与 \quad Q_i = P(G|E_i).$$

我们以质点第一次运动的情况为条件, 即以 $B = \{$ 质点第一次向左一单位 $\}$ 与 $\overline{B}$ 为分割, 用关于条件概率 $P(\cdot|E_i)$ 的全概公式可得

$$\begin{aligned}
Q_i &= P(G|E_i) \\
&= P(B|E_i)P(G|BE_i) + P(\overline{B}|E_i)P(G|\overline{B}E_i) \\
&= qQ_{i-1} + pQ_{i+1}, \quad i = 1, 2, \cdots, a-1.
\end{aligned}$$

将它变形为便于递推的形式

$$Q_{i+1} - Q_i = \frac{p}{q}(Q_i - Q_{i-1}), \quad i = 1, 2, \cdots, a-1. \tag{5.6}$$

(1) 当 $p = q = 1/2$ 时, 此式表明 $\{Q_i\}$ 组成一等差数列, 联合边值条件 $Q_0 = 0$ 及 $Q_a = 1$ 立得

$$Q_i = \frac{i}{a}, \quad i = 0, 1, \cdots, a.$$

(2) 当 $p \neq q$ 时, 反复用 (5.6) 式并注意 $Q_0 = 0$ 可导出

$$Q_{i+1} - Q_i = \left(\frac{q}{p}\right)^i Q_1, \quad i = 1, 2, \cdots, a-1. \tag{5.7}$$

上式对 $i = 1, 2, \cdots, a-1$ 求和并用 $Q_a = 1$ 可得

$$Q_1 = \left[\sum_{i=0}^{a-1}\left(\frac{q}{p}\right)^i\right]^{-1} = \frac{1 - q/p}{1 - (q/p)^a}.$$

再次将 (5.7) 式对 $i = 1, 2, \cdots, i-1$ 求和并将上式之 Q_1 代入便得到

$$Q_i = \frac{1 - (q/p)^i}{1 - (q/p)^a}, \quad i = 0, 1, \cdots, a.$$

总之，我们得到自 i 出发的随机游动终将被 a 吸收的概率

$$Q_i = \begin{cases} i/a, & p = q; \\ \frac{1-(q/p)^i}{1-(q/p)^a}, & p \neq q, \end{cases} \qquad i = 0, 1, \cdots a.$$

类似地可以算得，质点将在 0 处被吸收的概率为

$$P_i = \begin{cases} (a-i)/a, & p = q, \\ \frac{(q/p)^i-(q/p)^a}{1-(q/p)^a}, & p \neq q. \end{cases}$$

上述结果表明，对一切 i 总有 $P_i + Q_i = 1$. 就是说，两端带吸收壁 0 与 a 的随机游动终将被吸收的概率为 1, 从而质点在 1 至 $a-1$ 之间永远地徘徊下去的概率为 0.

这个例子在历史上以"输光问题"的形式最先出现. 假定甲有赌本 i 元，乙有 $a-i$ 元. 每赌一局，甲总以概率 p 赢乙 1 元，以概率 q 输给乙 1 元. 于是甲的赌本就是直线上的随机徘徊. 而甲的赌本到达 0 或 a 则表明甲或乙已输光. 这样， P_i 与 Q_i 分别为甲、乙二人的输光概率. 计算表明，乙输光的概率 Q_i 是甲的初始赌本数 i 的上升函数. 特别当 $p=q$ 时，即公平赌博时， Q_i 与 i 成正比. 例如甲有 10 元乙有 5 元，则 $p=q$ 时乙输光的概率 $Q_{10} = 2/3$. 但是，当 $p \neq q$ 时，情况将有较大变化. 仍以甲有 10 元乙有 5 元为例，如果取 $p = 0.4$, 即当每一局稍微对乙有利时，则乙输光的概率下降到 $Q_{10} = 0.13$.

1.5.4 习题

1. 假定生男孩或生女孩是等可能的. 在一个有 3 个孩子的家庭中，已知有男孩，求至少一个女孩的概率.

2. 掷 3 枚均匀的骰子，已知点数不同，求至少有一个 6 点的概率.

3. 袋中有 $2n-1$ 只白球与 $2n$ 只黑球，现任取出 n 只，发现它们是同色的，求同为黑色的概率.

4. 将 3 个不同的球任放入编号为 1,2,3,4 的 4 个盒，每球入各盒均等可能．现已知恰有 2 个空盒，求有球盒的最小编号为 k 的概率 $(k=1,2,3,4)$.

5. 举例说明对于任意事件 A,B,C, 下列等式不一定成立：

(1) $P(A|C)+P(\overline{A}|\overline{C})=1$;

(2) $P(A|C)+P(A|\overline{C})=1$;

(3) $P(A\bigcup B|C)=P(A|C)+P(B|C)$;

(4) $P(AB|C)=P(A|C)\cdot P(B|C)$.

试讨论使上列各式成立的条件.

6. 将分别写有字母 L, L, E, E, V 的 5 张卡片任排成一行，求恰好排成单词 LEVEL 的概率.

7. 将 4 个红球与 6 个黑球混放在袋中，现任意地一一摸出，直至红球全摸出为止．求恰好摸 k 次的概率.

8. 袋中有 r 个红球与 b 个黑球，现任取出一球，添加 s 个同色球一并放回．再从袋中任取出一球发现是红球，求第一次取出的是黑球的概率.

9. 假设一架坠毁的飞机掉在三个可能区域中的任何一个是等可能的．如果飞机坠落在区域 i 中 $(i=1,2,3)$, 则由于地理环境的影响，经过快速检查后发现其残骸的概率为 $\alpha_i(0<\alpha_i<1)$. 现快速检查区域 1 之后未发现残骸，求飞机坠落在区域 i 的条件概率.

10. 在电报通讯中，发射端发送“点”与“划”两种信号分别占 60% 与 40%. 由于随机干扰，接受端可能收到“点”,“划”与“不清”三种结果．假定发射“点”时，收到“点”、“划”,“不清”的概率依次是 0.7, 0.1, 0.2; 而发射“划”时收到上述三种结果的概率相应为 0, 0.9, 0.1. 分别计算收到结果为“不清”时，发出的信号为“点”,“划”的概率.

11. n 对夫妇任站成一列，求丈夫全在其妻子后面 (不一定相邻) 的概率 p_n.

12. 抛硬币 n 次，第 1 次抛出正面的概率为 c, 此后每次掷出与前次相同结果的概率为 $p\,(0\le p\le 1)$. 求第 n 次抛出正面的概率，并

讨论 $n \to \infty$ 时的极限.

13. 甲乙两袋各装有 1 个红球与 1 个黑球, 每次从两袋中各任取 1 球交换后放回. 求经 n 次交换后, 甲袋中包含 2 个红球, 1 红 1 黑, 2 个黑球的概率 p_n, q_n, r_n.

14. r 个人 $(r>1)$ 作传球游戏, 从某甲开始. 每次持球者均等可能地传给其余 $r-1$ 个人中的任一个. 求下列事件的概率: (1) 传了 n 次, 球没有回到过甲手中; (2) 传了 n $(n<r-1)$ 次, 没有人接到过两次球 (甲开始时持球算作已接球 1 次); (2) 第 n 次仍由甲传出.

15. 接连抛一枚均匀硬币, 直至第一次出现两个相连的正面为止. 求恰抛 n 次的概率. （提示: 记 $B_i=\{$ 第 i 次抛出正面 $\}$, 以 $\overline{B}_1, B_1B_2, B_1\overline{B}_2$ 为分割用全概公式. ）

1.6 事件的独立性

1.6.1 两个事件的独立性

同一随机试验中的各事件之间应当是相互有联系的. 这种联系反映在其中一个事件的发生对其它事件出现概率的影响. 就是说, 事件的无条件概率与已知另一事件发生后的条件概率一般是不同的. 但是在特殊情况下二者相同, 即一事件的发生与否对其它事件发生的概率无影响. 且看如下的简单例子.

例 6.1 袋中有 r 个红球与 b 个黑球, 现任意取出 2 球, 令 $R_i=\{$ 第 i 个是红球 $\}$. 则无论放回取球或不放回取球都有 $P(R_i)=r/(r+b), i=1,2$. 但是为算 $P(R_1R_2)$ 及 $P(R_2|R_1)$ 就必须分别讨论两种情形. 利用例 2.6 中已有结果 (2.7) 与 (2.6) 式可知, 对于有放回情形,

$$P(R_1R_2)=\frac{r^2}{(r+b)^2}, \quad P(R_2|R_1)=\frac{r}{r+b}=P(R_2).$$

而在不放回情形,

$$P(R_1R_2) = \frac{r(r-1)}{(r+b)(r+b-1)},$$

从而

$$P(R_2|R_1) = \frac{r-1}{r+b-1} \neq P(R_2).$$

得到这样的结果是很自然的. 对于不放回情形, 第一次取球的结果对第二次取球有影响. 而有放回时, 当然不存在这种影响. 事件之间这种互不影响各自发生概率的性质就是我们要介绍的独立性, 它是概率论中一个重要的基本概念.

定义 6.1 设 $(\Omega, \mathcal{F}, P)$ 是一个概率空间, 如果 $A, B \in \mathcal{F}$ 满足

$$P(AB) = P(A)P(B), \tag{6.1}$$

则称事件 A 与 B **相互独立**.

当 $P(B) > 0$ 时, 易证 A 与 B 独立的充分必要条件是

$$P(A|B) = P(A).$$

于是, 例 6.1 中放回情形的事件 A_1 与 A_2 是独立的. 但是人们不用此条件作为独立性的定义, 而宁愿用 (6.1) 式, 这是因为 (6.1) 式不排斥零概率事件, 而且式中事件 A 与 B 的地位对称, 反映了独立的"相互"性质. 下面是独立事件的又一例子.

例 6.2 向区间 $[0,1)$ 内随机地任投一点, 以事件 A 表示点落在子区间 $[0,1/2)$ 内, 而 B 表示点落在子区间 $[1/4,3/4)$ 内, 则 $P(A) = P(B) = 1/2$. 再由 $AB = [1/4,1/2)$ 可知 $P(AB) = 1/4 = P(A)P(B)$. 即 (6.1) 式成立, 事件 A 与 B 是相互独立的. 此例不象例 6.1 中的两个独立事件那么符合直观. 但是我们的出发点应当是数学定义 (6.1) 式, 而不能单凭直观.

应当指出，独立性显然是相对于概率 P 而言的. 在这个例子中，如果用 $\Omega=[0,1)$ 上的其它概率测度代替那里的勒贝格测度，则上述事件 A 与 B 就不一定独立了. 仍取 $\mathcal{F}$ 为 $[0,1)$ 中的波雷尔集类，考虑 Ω 中的二等分点列 $p_i=1/2^i, i=1,2,\cdots$. 则 $p_i>0$ 且 $\sum\limits_{i=1}^{\infty}p_i=1$. 对任何波雷尔集 $E\in\mathcal{F}$, 取 $\tilde{P}(E)=\sum\limits_{p_i\in E}p_i$. 易证 $(\Omega,\mathcal{F},\tilde{P})$ 为一概率空间. 相对于概率测度 $\tilde{P}$, 我们有 $\tilde{P}(A)=\sum\limits_{i=2}^{\infty}p_i=1/2, \tilde{P}(B)=p_1+p_2=3/4$, 而 $\tilde{P}(AB)=p_2=1/4\neq\tilde{P}(A)\tilde{P}(B)$. 故作为新的概率空间 $(\Omega,\mathcal{F},\tilde{P})$ 中的两个事件，上述 A 与 B 不独立.

由定义不难验证，概率为 0 或 1 的事件与任何其它事件都是独立的. 再者，若 A 与 B 独立，则 A 与 $\overline{B}$, $\overline{A}$ 与 B, $\overline{A}$ 与 $\overline{B}$ 都是相互独立的.

应注意区分这里介绍的“独立”与前面学过的“不相容”这两者的不同. “不相容”是说两个事件同时发生是不可能的，即它们没有公共的样本点，并不涉及概率测度. 不难验证，对于概率空间中的任意两个有正概率的事件，不相容必定是不独立的，独立也必定不是不相容的.

1.6.2 多个事件的独立性

定义 6.2 设 $(\Omega,\mathcal{F},P)$ 为一概率空间， $A_1,\cdots,A_n\in\mathcal{F}$. 如果对任意的 $s, 1\le s\le n$, 及任意的 $1\le i_1<i_2<\cdots<i_s\le n$ 都有

$$P(A_{i_1}A_{i_2}\cdots A_{i_s})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_s}), \qquad (6.2)$$

则称事件 $A_1,\cdots,A_n$ 相互独立.

注意 (6.2) 式实际上包含如下的 $2^n - n - 1$ 个式子:

$$\left.\begin{aligned} &P(A_{i_1}A_{i_2}) = P(A_{i_1})P(A_{i_2}),\ 1 \le i_1 < i_2 \le n,\\ &P(A_{i_1}A_{i_2}A_{i_3}) = P(A_{i_1})P(A_{i_2})P(A_{i_3}),\\ &\qquad\qquad 1 \le i_1 < i_2 < i_3 \le n,\\ &\cdots\qquad\cdots\qquad\cdots\\ &P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n). \end{aligned}\right\} \tag{6.3}$$

由上述定义立刻看出, 如果 $A_1, \cdots, A_n$ 这 n 个事件相互独立, 则其中任意 m 个 $(2 \le m \le n)$ 也相互独立, 特别, 每两个事件独立. 反过来说, 两两独立却不能推出整体独立. 我们以 $n = 3$ 为例, 此时 A_1, A_2, A_3 独立的定义是如下四个等式成立:

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3), \tag{6.4}$$

$$\left.\begin{aligned} P(A_1A_2) = P(A_1)P(A_2),\\ P(A_2A_3) = P(A_2)P(A_3),\\ P(A_3A_1) = P(A_3)P(A_1). \end{aligned}\right\} \tag{6.5}$$

下面举例说明 (6.4) 式不能推出 (6.5) 式, 同时 (6.5) 式也不蕴含 (6.4) 式.

例 6.3 (1) 设 $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, 样本点 ω_i 出现的概率为 p_i. 如果取 $p_i = 1/4, i = 1,2,3,4$. 而 $A_1 = \{\omega_1, \omega_2\}, A_2 = \{\omega_1, \omega_3\}, A_3 = \{\omega_1, \omega_4\}$. 则 $P(A_i) = 1/2, i = 1,2,3$. $P(A_1A_2) = P(A_2A_3) = P(A_1A_3) = p_1 = 1/4$. 故 (6.5) 式成立. 但 $P(A_1A_2A_3) = p_1 = 1/4 \ne P(A_1)P(A_2)P(A_3)$, (6.4) 式不成立.

(2) 设 $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, 对一切 i, 样本点 ω_i 出现的概率 $p_i = 1/6$. 如果取 $A_1 = \{\omega_1, \omega_2, \omega_3, \omega_4\}, A_2 = A_3 = \{\omega_4, \omega_5, \omega_6\}$. 则有 $P(A_1A_2A_3) = p_4 = 1/6 = P(A_1)P(A_2)P(A_3)$, 即 (6.4) 式成立. 但是 $P(A_2A_3) = P(A_2) = 1/2 \ne 1/4 = P(A_2)P(A_3)$. 故 (6.5) 式不成立.

最后我们给出任意多个事件相互独立的概念.

定义 6.3 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，每个 $t \in T$ 有 $A_t \in \mathcal{F}$. 如果对 T 的任有限子集 $\{t_1, t_2, \cdots, t_s\}$, 事件 $A_{t_1}, A_{t_2}, \cdots, A_{t_s}$ 相互独立，则称 $\{A_t : t \in T\}$ 为**独立事件族**.

根据这个定义可知，$\mathcal{F}$ 中事件序列 $\{A_n\}$ 为相互独立的充要条件是，任 $n \geq 1$ 有

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n).$$

1.6.3 独立性对概率计算的简化

在将概率论应用于实际生产与生活各领域的过程中，人们经常不言而喻地认定独立性条件满足. 即使对于系统中看上去有明显相依性的各个环节，也近似地假定它们是独立运行的. 之所以这样做，根本原因在于独立性假设可以大大简化有关概率的计算.

定理 6.1 如果 $A_1, A_2, \cdots, A_n$ 是某个概率空间中的相互独立事件，则乘法定理 (见定理 5.2) 可简化为

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n), \tag{6.6}$$

而加法定理 (见定理 4.2 之 9°) 也简化为

$$P(\bigcup_{k=1}^{n} A_k) = 1 - \prod_{k=1}^{n} [1 - P(A_k)]. \tag{6.7}$$

证明: (6.6) 式只不过是独立性定义 (6.3) 式的最后一式. 再由对偶原理有

$$\bigcup_{k=1}^{n} A_k = \overline{\bigcap_{k=1}^{n} \overline{A}_k},$$

但 $\overline{A}_1, \overline{A}_2, \cdots, \overline{A}_n$ 也相互独立 (参见本节习题 11). 至此 (6.7) 式可由 (6.6) 式推出.

下面是这个定理在系统的可靠性分析中的简单应用.

例 6.4 可靠性 系统(或组成系统的部件)正常工作的概率称作该系统(部件)的可靠性. 我们总假定系统中各部件能否正常工作是相互独立的. 设第 i 个部件的可靠性为 $p_i(0 < p_i < 1), i = 1, 2, \cdots, n$. 那么, 由定理 6.1 可知, 将这 n 个部件串联所组成的系统的可靠性为 $R_s = p_1 p_2 \cdots p_n$; 而将它们并联所组成系统的可靠性 $R_b = 1 - (1 - p_1)(1 - p_2) \cdots (1 - p_n)$. 对于其它较复杂的系统, 一般可以分解为若干串联或并联子系统, 而逐步求得其可靠性.

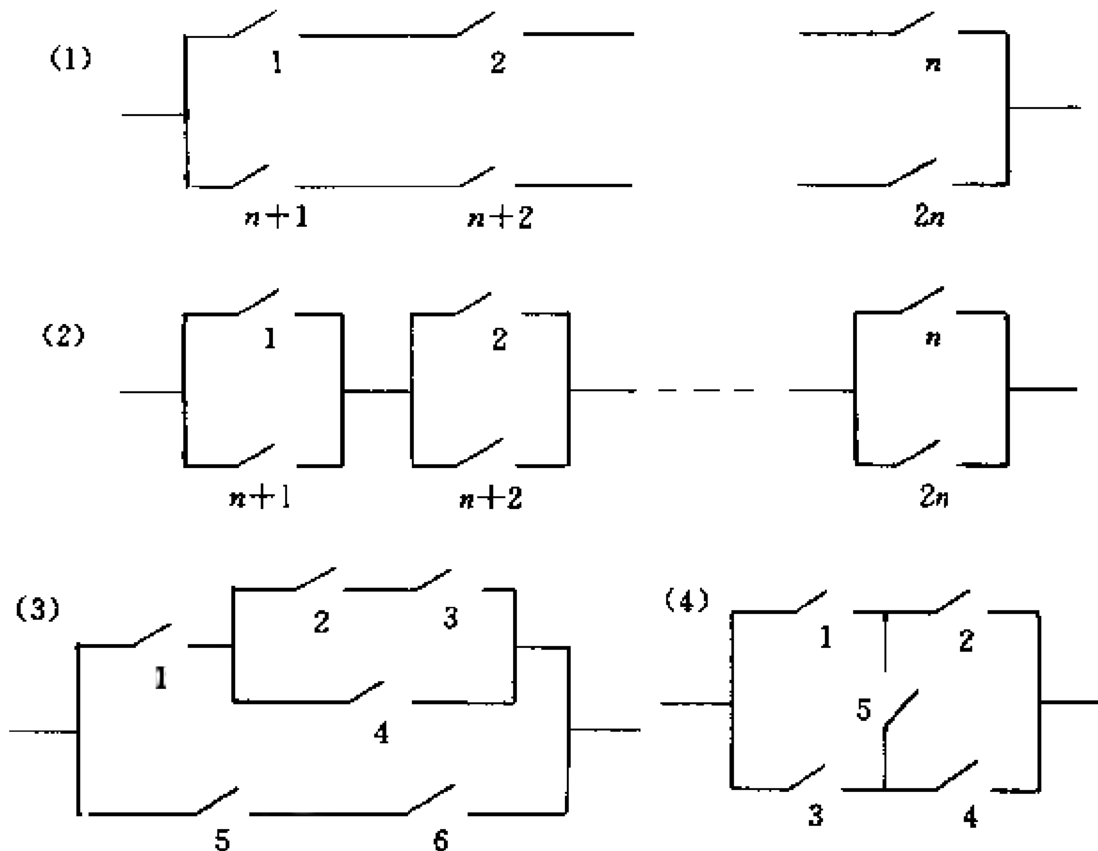


图 6.1 系统可靠性分析的例

假设图 6.1 中各开关接通的概率均为 $p(0 < p < 1)$. 试分别计算四个电路的两端为通路的概率 R_1, R_2, R_3, R_4.

解: 在电路 (1) 中, 开关 1 至 n 串联的接通概率为 p^n. 同样开关 $n+1$ 至 $2n$ 接通的概率也是 p^n. 它们再并联所得电路 (1) 的接通概率为 $R_1 = 1 - (1 - p^n)^2 = p^n(2 - p^n)$. 在电路 (2) 中, 对每个 $i = 1, 2, \cdots, n$, 开关 i 与 $(n + i)$ 并联的接通概率为 $p_i = p(2 - p)$, 这

n 个子电路再串联所得 (2) 的接通概率为 $R_2 = p^n(2-p)^n$. 易证, $n \geq 2$ 时有 $R_1 < R_2$. 这说明, 虽然电路 (1) 与 (2) 所用元件相同, 但采用后一种接法可提高系统的可靠性.

我们仍可用上述分解方法计算 R_3. 下面介绍另一种算法. 记 $A_i = \{$ 开关 i 接通 $\}$. 则不难看到, 电路 (3) 为连通的概率为

$$\begin{aligned}
R_3 &= P\left[(A_1A_2A_3)\bigcup(A_1A_4)\bigcup(A_5A_6)\right] \\
&= P(A_1A_2A_3) + P(A_1A_4) + P(A_5A_6) - P(A_1A_2A_3A_4) \\
&\quad -P(A_1A_4A_5A_6) - P(A_1A_2A_3A_5A_6) + \\
&\quad +P(A_1A_2A_3A_4A_5A_6) = 2p^2 + p^3 - 2p^4 - p^5 + p^6.
\end{aligned}$$

这里我们对 3 个事件 $A_1A_2A_3, A_1A_4, A_5A_6$ 运用了加法定理. 注意, 这 3 个事件不是两两不相容的, 故不能用有限可加性; 它们又不是相互独立的, 也不能用 (6.7) 式.

为求 $E = \{$ 电路 (4) 两端为通路 $\}$ 的概率, 我们以 A_5 及 $\overline{A}_5$ 为分割用全概公式, 则有

$$R_4 = P(E) = P(A_5)P(E|A_5) + P(\overline{A}_5)P(E|\overline{A}_5).$$

注意在开关 5 接通条件之下, (4) 相当于 $n = 2$ 时的电路 (2), 故 $P(E|A_5) = p^2(2-p)^2$. 同理, 在开关 5 断开条件之下, (4) 相当于 $n = 2$ 时的电路 (1). 从而 $P(E|\overline{A}_5) = p^2(2-p^2)$. 代入便得

$$\begin{aligned}
R_4 &= p \cdot p^2(2-p)^2 + (1-p) \cdot p^2(2-p^2) \\
&= 2p^2 + 3p^3 - 5p^4 + 2p^5.
\end{aligned}$$

可见全概公式在概率计算中有广泛的用途.

1.6.4 试验的独立性 *

所谓试验相互独立, 就是其中一试验所得到的结果, 对其它各试验取得其可能结果的概率都没有影响. 本节将给出较严格的数学描述.

先考虑两个随机试验. 假定第 i 个试验的概率空间为 $(\Omega_i, \mathcal{F}_i, P_i)$, $i=1,2$. 按照我们对独立性的理解, 两个试验的独立性应当叙述为:

$$\begin{array}{c}\text{对任何 } A_i \in \mathcal{F}_i, i=1,2, A_1 \text{ 与 } A_2 \text{ 同时}\\ \text{发生的概率等于它们各自概率之乘积.}\end{array} \tag{6.8}$$

但是这一命题至少有两点不妥. 首先, "A_1 与 A_2 同时发生" 应当是这两个事件的交, 但它们分别是两个样本空间 Ω_1, Ω_2 的子集合, 无法进行运算. 其次, 两个概率空间有各自的 P_1 与 P_2, 但此时涉及两个试验, 故命题中 "同时发生的概率" 既不能用 P_1 也不能用 P_2 来度量. 可见, 要严格刻画两试验的独立性, 必须适当地构造一个同时描述这两个试验的新的概率空间 $(\Omega, \mathcal{F}, P)$. 下面我们来说明, 描述两个独立试验的概率空间应当是 $(\Omega_1, \mathcal{F}_1, P_1)$ 与 $(\Omega_2, \mathcal{F}_2, P_2)$ 的笛卡尔乘积.

对任何 $A_i \in \mathcal{F}_i, i=1,2$, 定义它们的笛卡尔乘积

$$A_1 \times A_2 = \{(\omega_1, \omega_2) : \omega_1 \in A_1 \text{ 且 } \omega_2 \in A_2\}.$$

取 $\Omega = \Omega_1 \times \Omega_2$. 若 $A_i \in \mathcal{F}_i, i=1,2$, $\Omega_1 \times \Omega_2$ 中形如 $A_1 \times A_2$ 之的子集合称为可测矩形. 可测矩形的全体记为 $\mathcal{C}$. 然后取 $\mathcal{C}$ 生成的 σ 代数为 $\mathcal{F}_1$ 与 $\mathcal{F}_2$ 的笛卡尔乘积, 即 $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 = \sigma(\mathcal{C})$. 最后, 对于每个可测矩形 $A_1 \times A_2 \in \mathcal{C}$, 取

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \quad A_i \in \mathcal{F}_i, \quad i=1,2. \tag{6.9}$$

可以证明 (见参考书 [7], p.65), $\mathcal{C}$ 上的这个集函数 P 可唯一地扩张为 $\mathcal{F}_1 \times \mathcal{F}_2$ 上的概率测度, 称之为 P_1 与 P_2 的 乘积测度, 记为 $P = P_1 \times P_2$. 至此建立了乘积概率空间 $(\Omega, \mathcal{F}, P)$. 其中, 形如 $A_1 \times \Omega_2$ 的可测矩形由这样的样本点 (ω_1, ω_2) 组成: ω_1 限定在 A_1 中, 而 ω_2 在 Ω_2 中任取. 于是它代表 "第一个试验出现结果 A_1" 这种只涉及第 1 个试验的事件. 同理, 形如 $\Omega_1 \times A_2$ 的可测矩形代表 "第 2 个试验出现结果 A_2" 这个只涉及第二个试验的事件. 由 (6.9) 式可知

$$P(A_1 \times \Omega_1) = P_1(A_1), \quad P(\Omega_1 \times A_2) = P_2(A_2). \tag{6.10}$$

即上述只涉及一个试验的事件仍保持其在原概率空间中的概率值. 而用乘积测度来度量二者同时发生的概率时, 用 (6.9) 和 (6.10) 式可得

$$
\begin{aligned}
& P\{(A_1 \times \Omega_2) \bigcap (\Omega_1 \times A_2)\} = P(A_1 \times A_2) \\
= \;& P_1(A_1)P_2(A_2) = P(A_1 \times \Omega_2)P(\Omega_1 \times A_2). \qquad (6.11)
\end{aligned}
$$

这正是前面的命题 (6.8) 的严格数学表达. 可见将它们都放进乘积概率空间之后, 前面对 (6.8) 提出的两点不妥均不复存在.

至此我们可以说, 两个试验 $(\Omega_i, \mathcal{F}_i. P_i), i = 1,2$ 是否相互独立, 就是要看乘积样本空间 $\Omega_1 \times \Omega_2$ 上的概率是否是由 (6.9) 式确定的乘积测度. 就是说, 试验是否具有独立性取决于乘积空间上概率测度的选择.

为帮助读者 (特别是不熟悉乘积测度空间的读者) 理解试验的独立性, 我们结合一个简单的例子再作一说明.

考虑本节开头引进事件独立性概念的例 6.1. 从 r 个红球与 b 个白球中任取两球, 当时我们是用一个概率空间 $(\Omega, \mathcal{F}, P)$ 来描述这个试验. 例如对于有放回情形, $\Omega = \{(\omega_1, \omega_2)\}, \omega_1$ 表示第 1 次取球的结果, 共有 $r + b$ 个可能, ω_2 表示第 2 次取球的结果, 也有 $r + b$ 个可能. 从而 $n(\Omega) = (r + b)^2$. 而对事件 $A_i = \{$ 第 i 个是红球 $\}$ 有 $n(A_i) = r(r + b)$. 则 $P(A_i) = r/(r + b), i = 1, 2$. 另一方面 $n(A_1A_2) = r^2$, 故 $P(A_1A_2) = r^2/(r + b)^2 = P(A_1)P(A_2)$, 两事件 A_1 与 A_2 独立. 其实, 如果将两次取球分开看作两个试验, 对 $i = 1, 2$, 取 $\Omega_i = \{\omega_i\}$, 其中含有 $r + b$ 个等可能的样本点, 故 $P_i(A_i) = r/(r + b)$. 这样就构造出两个试验各自的概率空间 $(\Omega_i, \mathcal{F}_i, P_i)$. 原来那个描述两次取球的概率空间 $(\Omega, \mathcal{F}, P)$ 正是它们的笛卡尔乘积. 就是说, 对放回取球情形, 两次取球 (试验) 是相互独立的. 有的读者可能会认为, 乘积空间上的测度就只能是乘积测度. 就象算术中矩形面积等于长乘宽一样自然. 其实不然. 即使对欧氏空间中的图形, 除了面积与长度 (勒贝格测度) 之外, 尚有许多其它的度量. 回到例 6.1, 在不放回取球情形, 虽然仍然可以用乘积样本空间 $\Omega = \{(\omega_1, \omega_2)\}$ 表示两次取球的结果, 但是其中的 $(r + b)^2$ 个样本点不再是等可能的了. 例如, 使

$\omega_1 = \omega_2$ 的样本点是不可能出现的. 就是说, 在描述不放回取球两次的概率空间中, 概率 P 已不再是乘积测度, 从而两次取球试验不独立.

最后我们给出 n 个试验相互独立的一般定义. 所用记号与术语的含义, 与前面 $n = 2$ 情形没有本质的区别.

定义 6.4 设有 n 个随机试验, 第 i 个试验的概率空间为 $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2, \cdots, n$. 代表这 n 个试验的乘积样本空间 $\Omega = \Omega_1 \times \cdots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_n = \sigma(\mathcal{C})$ (其中 $\mathcal{C}$ 为形如 $B_1 \times \cdots \times B_n, B_i \in \mathcal{F}_i$ 的可测矩形全体). 如果 $(\Omega, \mathcal{F})$ 上的概率测度 P 是 $P_1, \cdots, P_n$ 的乘积测度, 即对任何 $B_1 \times \cdots \times B_n \in \mathcal{C}$ 满足

$$P(B_1 \times \cdots \times B_n) = P_1(B_1) \cdot \cdots \cdot P_n(B_n), \tag{6.12}$$

则称这 n 个试验是相互独立的.

如果再设 $\Omega_i \equiv \Omega_0, \mathcal{F}_i \equiv \mathcal{F}_0, P_i \equiv P_0, i = 1, 2, \cdots, n$, 即 n 个试验有相同的概率空间, 则称它们为 n 个独立重复试验.

1.6.5 习题

1. 假定生男孩生女孩是等可能的, 试证在有 3 个孩子的家庭中, 事件 $E = \{$ 此家庭至多有 1 个女孩 $\}$ 与 $F = \{$ 家庭中有男孩也有女孩 $\}$ 相互独立. 如果考虑有 2 个孩子的家庭呢?

2. 掷 2 枚均匀的骰子, 设 $A = \{$ 第 1 枚为 4 点 $\}$, $B = \{$ 点数和为 6$\}$, $C = \{$ 点数和为 7$\}$. 试讨论每两个事件间的独立性.

3. 设事件 A 与 B 独立, 且两事件中仅 A 发生, 仅 B 发生的概率都是 1/4. 求 $P(A)$ 与 $P(B)$.

4. 设 A, B, C 三个事件独立, 求证 $A \bigcup B, A - B$ 都与 C 独立.

5. 掷 2 枚均匀骰子, 记 $A = \{$ 点数和为奇数 $\}$, $B = \{$ 第 1 枚为奇数点 $\}$, $C = \{$ 第 2 枚为偶数点 $\}$. 求证 A, B, C 两两独立但不整体独立.

6. 在习题 1.5.4 的第 11 题中，记 $A_i = \{$ 第 i 个丈夫站在其妻子后面 $\}$, 证明事件 $A_1, \cdots, A_n$ 相互独立.

7. 对同一目标进行 3 次独立的射击，其命中概率依次为 0.4, 0.5, 0.7. 试求下列事件的概率：(1) 3 次中恰有一次命中，(2)3 次中至少 1 次命中.

8. 当开关 K_1 断开，或开关 K_2 与 K_3 同时断开时电路断开. 设 K_1, K_2, K_3 断开的概率依次是 0.4, 0.5, 0.7, 且各开关相互独立. 求电路为断开的概率.

9. 假设每个人血清中含肝炎病毒的概率为 0.004, 且各个人血清是否含肝炎病毒相互独立. 求 100 个人血清混合后含肝炎病毒的概率.

10. 无线电监测站负责监测 n 个目标. 假定在监测过程中第 i 个目标消失的概率为 p_i, 且各目标是否消失相互独立. 求下列事件的概率：(1) 监测过程中没有目标消失；(2) 至少一个目标消失；(3) 不多于一个目标消失.

11. 试证事件 $A_1, \cdots, A_n$ 相互独立的充分必要条件是，对每个 $\widehat{A}_k = A_k$ 或 $\overline{A}_k (k = 1, \cdots, n)$ 总有

$$P(\widehat{A}_1 \cdots \widehat{A}_n) = P(\widehat{A}_1) \cdots P(\widehat{A}_n).$$

12. 假定 $0 < P(B) < 1$, 试证明事件 A 与 B 独立的一个充分必要条件是 $P(A|B) = P(A|\overline{B})$.

13. 如果事件 A, B, C 满足习题 1.5.4 中 5 题的 (4) 式，则称 A 与 B 关于 C **条件独立**. 试证明

(1) 当 $P(BC)>0$ 时，上述条件独立的充分必要条件是 $P(A|BC) = P(A|C)$.

(2) 举例说明 “独立” 与 “条件独立” 两者没有蕴含关系.

14. 假定事件 A, B, C 两两独立，且满足

$$ABC = \emptyset \qquad \text{及} \qquad P(A) = P(B) = P(C) = x.$$

试证 x 的最大值为 1/2. [**提示**：研究事件 $A \bigcup B \bigcup C$ 发生概率的上下界.]

第二章 随机变量

2.1 随机变量及其分布

2.1.1 定义与等价条件

随机试验的结果经常是数量. 例如, 掷一枚骰子所得点数, 记录电话呼唤次数, 预报明天的最高气温等等, 所得的可能结果都是数量. 有的随机试验的结果虽不是数, 但可以将它数量化. 例如抛一枚硬币, 可能结果 ω_1 =“正面” 与 ω_2 =“反面” 不是数. 如果约定以 “1” 代表正面, 以 “0” 代表反面, 即取

$$\xi(\omega)=\begin{cases}1, & \text{当}\ \ \omega=\omega_1,\\ \\ 0, & \text{当}\ \ \omega=\omega_2,\end{cases} \tag{1.1}$$

则这个有两个可能值的数量 $\xi(\omega)$ 代表了抛一枚硬币这一试验的结果.

作为随机试验的结果, 这些数量与以往用来表示时间, 位移等的变量有很大不同, 这就是其取值的变化情况取决于随机试验的结果, 因而是不能完全预言的. 这种随机地取值的变量就是随机变量. 先看两个简单的例子.

例 1.1 有朋自远方来. 他可能乘船, 乘火车到达, 也可能乘飞机到达. 如记 ω_1 = 乘船, ω_2= 乘火车, ω_3= 乘飞机, 这就是以 $\Omega=\{\omega_1,\omega_2,\omega_3\}$ 为样本空间的随机试验. 现假定客人将采取上述三种旅行方式的概率依次为 1/4, 1/2, 1/4, 于是 Ω 上的概率测度 P 也取定, 从而建立了此随机试验的概率空间. 现考虑这位客人的旅费, 这是一个随旅行方式 (即样本点 ω) 而变化的变量, 假定乘船、火车

与乘飞机的票价分别为 100, 200, 300 元. 则所需旅费就是 Ω 上定义的如下实值函数

$$\xi(\omega)=\begin{cases}100, & 若\ \omega=\omega_1,\\ 200, & 若\ \omega=\omega_2,\\ 300, & 若\ \omega=\omega_3.\end{cases} \tag{1.2}$$

当然, 我们不仅要关心旅费 $\xi(\omega)$ 的可能值, 还要关心它取各个可能值的概率. 由前面的假设可知

$$P\{\xi=100\}=P\{\xi=300\}=1/4, \qquad P\{\xi=200\}=1/2.$$

例 1.2 某人一觉醒来, 打开收音机对表. 求他至少等待 50 分钟的概率.

这是一个一维几何概率问题. 我们以 "分" 作为时间单位, 则 $\Omega=[0,60]$, 所关心的事件 $A=\{$ 至少等待 50 分钟 $\}=[0,10)$. 故 $P(A)=m(A)/m(\Omega)=1/6$. 此试验的可能结果本身就是数量, 于是醒来时间 $\xi(\omega)=\omega$, $\omega\in[0,60]$, 就是描述此试验结果的随机变量. 由几何概率的定义可知, 对 $x\in[0,60]$ 有

$$P\{\xi(\omega)<x\}=x/60, \tag{1.3}$$

特别取 $x=10$ 就是所求概率 $P(A)$.

在后一个例子中, $\xi(\omega)$ 可能值不再是有限多个, 而是遍布一个区间. 同时由于单点集的勒贝格测度 (即长度) 均为 0. 即恒有 $P\{\xi(\omega)=x\}=0$, 故不能象例 1.1 那样逐个给出 $\xi(\omega)$ 取各可能值的概率, 而是采取 (1.3) 式的形式给出 $\xi(\omega)$ 取值变化的概率规律. 从另一方面说, 如果对一切 x, 形如 $\{\xi(\omega)<x\}$ 这种事件的概率已经给出, 那么象例 1.1 中的 $\xi(\omega)$ 取各指定值的概率也就确定了. 这一点将在后面详加叙述. 于是, 采取给出一切形如 $\{\xi(\omega)<x\}$ 的事件概率的方法更广泛, 更适用. 当然, 首先应要求这种集合确是随机事件. 至此我们可以叙述随机变量的严格定义.

定义 1.1 设 $(\Omega, \mathcal{F}, P)$ 为概率空间，$\xi = \xi(\omega)$ 为 Ω 上定义的实值函数. 如果有

$$\{\omega : \xi(\omega) < x\} \in \mathcal{F}, \qquad \text{任} \quad x \in \Re. \tag{1.4}$$

则称 $\xi(\omega)$ 为**随机变量**.

此定义表明，随机变量就是 Ω 上关于 $\mathcal{F}$ 可测的实值函数. 今后时常省略 ω，而将 $\xi(\omega), \{\omega : \xi(\omega) < x\}$ 等简写为 $\xi, \{\xi < x\}$ 等.

作为样本空间 Ω 到 $\Re = (-\infty, +\infty)$ 的点变换 $\xi(\omega)$，可以引出将 $\Re$ 的子集 B 变为 Ω 的子集之逆变换:

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\}, \qquad \text{任} \quad B \subset \Re.$$

以下是将要用到的关于逆变换 ξ^{-1} 的性质.

定理 1.1 逆变换 ξ^{-1} 满足

1° $\xi^{-1}(\overline{B}) = \overline{[\xi^{-1}(B)]}$;

2° $\xi^{-1}(\bigcup_n B_n) = \bigcup_n \xi^{-1}(B_n)$.

证明: 仅以 2° 为例. 一方面对于任意的 n，若 $\xi \in B_n$ 则 $\xi \in \bigcup_n B_n$，故 $\xi^{-1}(B_n) \subset \xi^{-1}(\bigcup_n B_n)$. 从而有 $\bigcup_n \xi^{-1}(B_n) \subset \xi^{-1}(\bigcup_n B_n)$. 另一方面，由 $\xi \in \bigcup_n B_n$ 知存在 n 使 $\xi \in B_n$，从而 $\xi^{-1}(\bigcup_n B_n) \subset \bigcup_n \xi^{-1}(B_n)$. 2° 于是得证.

以下给出随机变量定义 (1.4) 式形式上更强的等价形式.

定理 1.2 概率空间 $(\Omega, \mathcal{F}, P)$ 上实值函数为随机变量的充分必要条件是

$$\xi^{-1}(B) \in \mathcal{F}, \qquad \text{任} \quad B \in \mathcal{B}. \tag{1.5}$$

这里的 $\mathcal{B}$ 是 $\Re$ 上的波雷尔集类，即由半直线全体组成的类 $\mathcal{P} = \{(-\infty, x) : x \in \Re\}$ 所生成的 σ 代数.

证明: 取 (1.5) 式中的 $B=(-\infty,x)\in\mathcal{B}$ 可得充分性. 为证其必要性, 考虑 $\Re$ 的子集类

$$\mathcal{E}=\{B\subset\Re:\xi^{-1}(B)\in\mathcal{F}\}.$$

则由 (1.4) 式知 $\mathcal{E}\supset\mathcal{P}$. 又由定理 1.1 知 $\mathcal{E}$ 为 σ 代数. 故 $\mathcal{E}\supset\mathcal{B}=\sigma(\mathcal{P})$. 此即 (1.5) 式成立. 证完.

考虑随机变量 ξ 的函数 $\eta=f(\xi)$. 由定理 1.2 知, 为要 η 仍为随机变量, 必须且只需对任意 $B\in\mathcal{B}$ 有 $\eta^{-1}(B)\in\mathcal{F}$. 但 $\eta^{-1}(B)=\{\omega:f[\xi(\omega)]\in B\}=\{\omega:\xi(\omega)\in f^{-1}(B)\}$. 故只要函数 f 满足 "对任何 $B\in\mathcal{B}$ 有 $f^{-1}(B)\in\mathcal{B}$" 即可. 满足这样要求的 f 称为**波雷尔函数**. 于是我们得到

定理 1.3 随机变量的波雷尔函数仍为随机变量.

常用的函数, 例如 $\Re$ 上分段连续函数, 均为波雷尔函数. 定理 1.3 在一定意义下的逆仍成立. 两者合起来就是著名的**杜勃复合函数定理**. (见参考书目 [7],19 页定理 17.)

2.1.2 随机变量的结构 *

设 $(\Omega,\mathcal{F},P)$ 为一概率空间. 对任何 $E\in\mathcal{F}$, 取事件 E 的示性函数 $\chi_E(\omega)=1$, 若 $\omega\in E$; $=0$, 否则. 易证 $\chi_E(\omega)$ 是随机变量. 这样, $\mathcal{F}$ 中的每一个事件都可以用其示性函数这个随机变量来代表.

下考虑样本空间 Ω 的任一有限分割 $\{A_1,A_2,\cdots,A_n\}$, 满足 $A_i\in\mathcal{F}\ (i=1,2,\cdots,n)$, $A_iA_j=\emptyset(i\neq j)$ 及 $\bigcup\limits_{i=1}^{n}A_i=\Omega$. 对任意 n 个不同的实数 $a_1,a_2,\cdots,a_n$ 取

$$\xi(\omega)=\sum_{i=1}^{n}a_i\chi_{A_i}(\omega). \tag{1.6}$$

作为 n 个示性函数的线性组合, $\xi(\omega)$ 仍为随机变量. 称形如 (1.6) 的 $\xi(\omega)$ 为**简单随机变量**.

对于 $(\Omega,\mathcal{F},P)$ 上任意非负随机变量 ξ 及自然数 n, 将 Ω 按 $\xi(\omega)$ 的取值进行分割. 即令

$$A_k=\left\{\omega:\frac{k}{2^n}\le\xi(\omega)<\frac{k+1}{2^n}\right\},\qquad k=0,1,\cdots,n2^n-1.$$

$$A_{n2^n}=\{\omega:\xi(\omega)\ge n\}. \tag{1.7}$$

不难验证 $\{A_0,A_1,\cdots,A_{n2^n}\}$ 满足前述 Ω 分割的条件. 从而

$$\xi_n(\omega)=\sum_{k=0}^{n2^n}\frac{k}{2^n}\chi_{A_k}(\omega) \tag{1.8}$$

为简单随机变量. 随机变量序列 $\{\xi_n\}$ 满足

$$0\le\xi_1(\omega)\le\cdots\le\xi_n(\omega)\to\xi(\omega). \tag{1.9}$$

我们总是运用 $\vee$ 与 $\wedge$ 表示取大与取小运算, 即 $a\vee b=\max(a,b)$, 而 $a\wedge b=\min(a,b)$. 对一般的随机变量 ξ, 令 $\xi^+=\xi\bigvee 0,\xi^-=(-\xi)\bigvee 0$. 则有 $\xi=\xi^+-\xi^-$, 且 ξ^+ 与 ξ^- 都是非负随机变量. 将 ξ^+,ξ^- 都按 (1.7) 至 (1.9) 表为简单随机变量列的极限, 则两个简单随机变量列之差的极限就是随机变量 ξ. 至此我们得到随机变量的结构定理.

定理 1.4 $(\Omega,\mathcal{F},P)$ 上实值函数 $\xi(\omega)$ 为随机变量的充分必要条件是, 存在简单随机变量序列 $\{\xi_n,n\ge1\}$ 使

$$\lim_n\xi_n(\omega)=\xi(\omega),\qquad 任\ \ \omega\in\Omega. \tag{1.10}$$

而且当 ξ 非负时, 还可取 $\{\xi_n\}$ 为非负简单随机变量的非降序列.

定理中条件充分性的证明留作习题.

2.1.3 分布与分布函数

在随机变量的数学定义中, 只涉及到样本空间 Ω 及其上事件 σ 代数 $\mathcal{F}$, 肯定了一切形如 $\xi^{-1}(B)$ 的 Ω 的子集在 $\mathcal{F}$ 中, 并没有涉及

概率测度 P. 但是，既然 $\xi^{-1}(B)$ 是随机事件，它就有相应的概率，取

$$\mathbf{F}(B) = P\{\xi^{-1}(B)\} = P\{\xi \in B\}, \qquad 任 \quad B \in \mathcal{B}. \tag{1.11}$$

定理 1.5 设 ξ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，则由 (1.11) 式定义的 $\mathcal{B}$ 上的集函数 $\mathbf{F}$ 是一个概率测度.

证明: 由 P 非负立得 $\mathbf{F}$ 的非负性. 再由 $\xi^{-1}(\Re) = \Omega$ 及 P 的规范性得 $\mathbf{F}(\Re) = 1$. 最后，设 $\{B_n\}$ 为 $\mathcal{B}$ 中不交集列. 一方面由定理 1.1 知 $\xi^{-1}(\bigcup_n B_n) = \bigcup_n [\xi^{-1}(B_n)]$, 另一方面，因 $i \neq j$ 时 $B_iB_j = \emptyset$, 故有 $\emptyset = \xi^{-1}(\emptyset) = \xi^{-1}(B_iB_j) = \xi^{-1}(B_i)\xi^{-1}(B_j)$. 这说明事件列 $\{\xi^{-1}(B_n)\}$ 是两两不相容的. 用 P 的可列可加性便得

$$\begin{aligned}\mathbf{F}(\bigcup_n B_n) &= P\{\xi^{-1}(\bigcup_n B_n)\} = P\{\bigcup_n [\xi^{-1}(B_n)]\} \\ &= \sum_n P\{\xi^{-1}(B_n)\} = \sum_n \mathbf{F}(B_n),\end{aligned}$$

即 $\mathbf{F}$ 有可列可加性. 于是得证 $\mathbf{F}$ 为 $\mathcal{B}$ 上的概率测度.

定义 1.2 称由 (1.11) 式定义的 $\mathcal{B}$ 上概率测度 $\mathbf{F}$ 为随机变量 ξ 的**概率分布**, 简称为**分布** 或**概率律**.

至此，概率空间 $(\Omega, \mathcal{F}, P)$ 上每个随机变量 ξ, 都在其值域 $\Re$ 上引进另一个概率空间 $(\Re, \mathcal{B}, \mathbf{F})$, 称之为 ξ 的相空间. 注意相空间中测度 $\mathbf{F}$ 与 ξ 有关.

将引进 ξ 的分布 $\mathbf{F}$ 的 (1.11) 式倒转来看，对于任何 $B \in \mathcal{B}$, 随机变量 ξ 落入 B 中的概率可通过计算 B 的测度 $\mathbf{F}(B)$ 得出. 这就是说，概率分布 $\mathbf{F}$ 完全刻画了 ξ 取值的概率规律. 下面我们给出描述这一规律的更简捷方便的工具.

将 $\mathcal{B}$ 上测度 $\mathbf{F}$ 仅局限于集类 $\mathcal{P} = \{(-\infty, x) : x \in \Re\}$ 上，由于 $\mathcal{P}$ 中每条半直线被它的右端点 x 所决定，于是集函数 $\mathbf{F}$ 就化为 $\Re$ 上的点函数.

定义 1.3 设 ξ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，称

$$F(x) = \mathbf{F}\{(-\infty, x)\} = P\{\xi < x\}, \qquad x \in \Re, \tag{1.12}$$

为随机变量 ξ 的**分布函数**.

容易得到，例 1.1 中旅费 ξ 的分布函数为

$$F(x) = \begin{cases} 0, & x \le 100, \\ 1/4, & 100 < x \le 200, \\ 3/4, & 200 < x \le 300, \\ 1, & 300 < x. \end{cases}$$

这是一个阶梯函数，它在 $x = 100, 200, 300$ 处依次有跃度为 $1/4$, $1/2$, $1/4$ 的跳跃. 其图形见图 1.1 的左图.

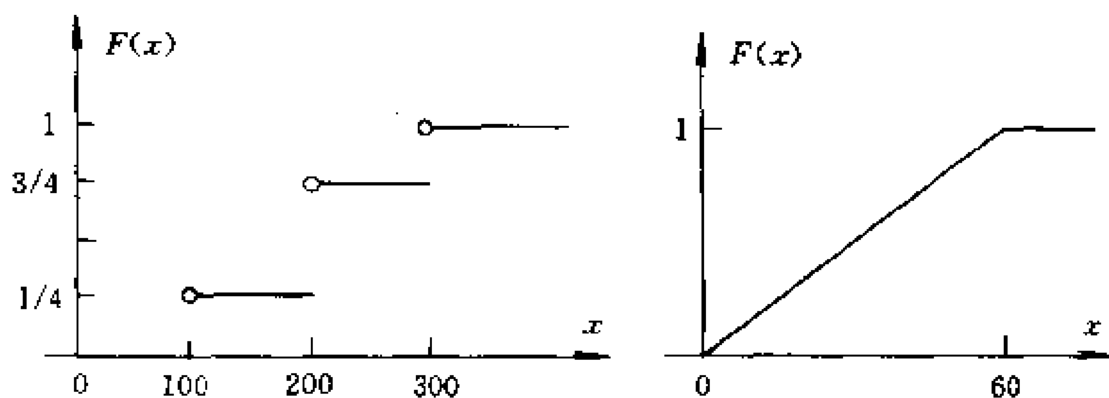


图 1.1　两种分布函数的图形

对于例 1.2 中醒来时刻 ξ，注意到 (1.3) 式及 $x \le 0$ 时 $\{\xi < x\} = \varnothing$, $x \ge 60$ 时 $\{\xi < x\} = \Omega$, 可得到此 ξ 的分布函数为

$$F(x) = \begin{cases} 0, & x \le 0, \\ x/60, & 0 < x < 60, \\ 1, & 60 \le x. \end{cases}$$

这是一个连续函数，其图形见图 1.1 的右图.

如下定理给出分布函数的一般性质.

定理 1.6 分布函数 $F(x)$ 具有

1° **单调非降性**: 任 $x_1 < x_2$ 有 $F(x_1) \le F(x_2)$;

2° **左连续性**: 任 $x_0 \in \Re$ 有 $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$;

3° **规范性**:

$F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0,\ F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1.$

证明: 首先，当 $x_1 < x_2$ 时有 $\{\xi < x_1\} \subset \{\xi < x_2\}$, 由概率的单调性可得 $F(x_1) \le F(x_2)$. 其次，对任何 $x_n \uparrow x_0 \in \Re$, 事件列 $A_n = \{\xi < x_n\}$ 为单调上升的，其极限为 $\bigcup\limits_n A_n = \{\xi < x_0\}$, 用概率的下连续性可得 $\lim\limits_n F(x_n) = F(x_0)$, 即 $F(x)$ 在 $\Re$ 上左连续. 最后，注意到事件列 $\{\xi < -n\} \downarrow \varnothing$, $\{\xi < n\} \uparrow \Omega$, 利用概率的上，下连续性及规范性，便得证 $F(-\infty) = 0$ 与 $F(+\infty) = 1$. 证完.

定理中的性质 $1°, 2°, 3°$ 是分布函数的本征性质. 就是说，每一个随机变量的分布函数必有这三条性质. 反过来，$\Re$ 上任何具备这三条性质的函数 $F(x)$ 都必定是某个随机变量的分布函数. 逆命题证明的一个关键是要证明满足 $1°, 2°, 3°$ 的 $F(x)$ 可唯一决定 $\mathcal{B}$ 上的概率测度 $\mathbf{F}$. 下面做一简要说明.

在定义 1.1 中给出随机变量 ξ 的数学定义之后，可以直接根据 (1.4) 式给出由 (1.12) 式定义的 ξ 的分布函数 $F(x)$, 并证明 $F(x)$ 有性质 $1°, 2°, 3°$. 视 $F(x)$ 为 $\mathcal{P}$ 上的集函数 $\mathbf{F}\{(-\infty, x)\} = F(x)$, 然后将 $\mathbf{F}$ 逐步扩张到更大的集类上去. 比如对区间 $[a, b)$ 可取 $\mathbf{F}\{[a, b)\} = F(b) - F(a)$, 对不交区间的有限并 $A = \bigcup\limits_{i=1}^{n} [a_i, b_i)$, 取 $\mathbf{F}(A) = \sum\limits_{i=1}^{n} \mathbf{F}\{[a_i, b_i)\}$, 等等，不再一一列举. 有测度扩张定理 (见参考书目 [7] p.37) 保证，此 $\mathbf{F}$ 可唯一地扩张到 $\mathcal{B}$ 上，成为 $\mathcal{B}$ 上的概率测度. 称之为分布函数 $F(x)$ 所引出的勒贝格—斯蒂尔吉斯测度. 实际上这个 $\mathbf{F}$ 正好是我们前面用 (1.11) 式引进的概率分布.

既然分布函数 $F(x)$ 可完全决定随机变量 ξ 的分布，于是它就完全描述了 ξ 的概率规律，即有关 ξ 的概率问题都可以借助 $F(x)$ 来回答. 例如对任意 $a < b$ 有

$$P\{\xi \in [a,b)\} = P\{\xi < b\} - P\{\xi < a\} = F(b) - F(a). \qquad (1.13)$$

又如，对任意 $a \in \Re$,

$$P\{\xi = a\} = \lim_n P\{\xi \in [a, a+\frac{1}{n})\} = F(a+0) - F(a), \qquad (1.14)$$

即 ξ 取任何指定值的概率等于其分布函数在该点上的跃度. 特别，ξ 取值于 $F(x)$ 的连续点上的概率为 0.

一般地，可将描述 ξ 取值概率规律的 (1.11) 式写为

$$P\{\xi \in B\} = \mathbf{F}(B) = \int_B dF(x), \qquad 任 \ B \in \mathcal{B}. \qquad (1.15)$$

这是一个在理论上及应用中都十分重要的公式. 此式的右方是关于分布函数 $F(x)$ 的勒贝格 - 斯蒂尔吉斯积分. 有关这种积分的知识可以在一般的实变函数论教科书中查到. 对这些内容不太熟悉的读者，可把 (1.15) 式当做诸如前面的 (1.13),(1.14) 以及后面的 (1.17),(1.22) 等特殊形式的一般记法.

2.1.4 离散型与连续型

离散型与连续型随机变量是两种最重要的特殊类型的随机变量. 本节开头的例 1.1 与例 1.2 正是这两种类型的例子，它们的分布函数分别是纯跳跃函数和绝对连续函数.

定义 1.4 设 ξ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，如果存在数列 $\{x_k\}$ 及 $\{p_k\}$, 满足 1° $p_k \ge 0, 2^\circ$ $\sum_k p_k = 1$ 使得

$$P\{\xi = x_k\} = p_k, \qquad k = 1, 2, \cdots, \qquad (1.16)$$

则称此随机变量 ξ(及其概率分布) 为**离散型的**, 而称由这两个数列组成的矩阵

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

为 ξ 的**密度阵**, 简称为**密度**.

由定义可知, ξ 取数列 $\{x_k\}$ 以外值的概率为 0. 但我们时常说成 "ξ 只取 $\{x_k\}$ 这至多可列个值". 这种说法 (也包括 "非负随机变量" 等类似提法) 实际上可能有零概率的偏差. 由于我们仅关心有关 ξ 的概率问题, 故总是允许这类偏差. 注意只有有限个可能值 $x_1, \cdots, x_n$ 的离散型随机变量正好是我们在 2.1.2 节引进的简单随机变量 $\xi(\omega) = \sum_{i=1}^{n} a_i \chi_{A_i}(\omega)$, 此处 $a_i = x_i, A_i = \{\xi = x_i\} \in \mathcal{F}, i = 1, \cdots, n$.

离散型随机变量 ξ 的密度阵完全决定了由 (1.11) 式定义的 ξ 的概率分布 $\mathbf{F}$. 事实上我们有

$$\mathbf{F}(B) = P\{\xi \in B\} = \sum_{k: x_k \in B} p_k, \qquad 任\quad B \in \mathcal{B}. \tag{1.17}$$

取上式中 $B = (-\infty, x)$, 可得 ξ 分布函数

$$F(x) = P\{\xi < x\} = \sum_{k: x_k < x} p_k, \qquad 任\quad x \in \Re. \tag{1.18}$$

易见, 这是一个纯跳跃函数: 它在 ξ 的每个可能值 x_k 上有跃度 p_k, 在每个不含诸 x_k 的区间上恒取常值. 关于这种分布函数的勒贝格 – 斯蒂尔吉斯积分恰好是 (1.17) 式右端的和式. 就是说, (1.17) 就是 (1.15) 式的离散型版本.

按照我们的定义, 只有一个可能值的常数 c 是离散型随机变量, 其密度阵为 $\begin{pmatrix} c \\ 1 \end{pmatrix}$. 称之为**退化分布**. 而抛一枚均匀硬币得到的随机变量 (1.1) 式具有两点分布 $\begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$.

我们从例 1.2 中的醒来时刻 ξ 入手讨论随机变量的另一重要类型. 将此 ξ 的分布函数求导可得

$$F'(x) = p(x) = \frac{1}{60}, \qquad 0 < x < 60.$$

这里及此后我们总约定: 凡是分段函数, 在已经注出的定义域之外其值恒为零. 比如上式的 $p(x)$ 在区间 $(0, 60)$ 之外的值为 0. 不再一一说明.

于是 ξ 的分布函数可写为

$$F(x) = \int_{-\infty}^{x} p(t)\, dt, \qquad x \in \Re. \tag{1.19}$$

注意此式与离散型分布函数的表达式 (1.18) 式非常相似. 那里是对一切小于 x 的可能值 x_k 将 p_k 求和, 这里是对一切小于 x 的可能值 t 将 $p(t)$ 积分.

定义 1.5 设 ξ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量. 如果存在函数 $p(x)$, 满足

$$1^\circ\ p(x) \geq 0, \qquad 2^\circ \int_{-\infty}^{+\infty} p(x)\, dx = 1, \tag{1.20}$$

使得 ξ 的分布函数 $F(x)$ 可表为 (1.19) 式的形式, 则称 ξ(及其概率分布) 为**连续型**的, 称 $p(x)$ 为 ξ 的**密度函数**, 简称为**密度**.

能够表为 (1.19) 式变上限积分的函数 $F(x)$ 在分析中称作绝对连续函数. 绝对连续函数必为连续函数. 于是用 (1.14) 式可得, 连续型随机变量取任何指定值 a 的概率 $P\{\xi = a\} = F(a+0) - F(a) = 0$.

由于在至多可列个点上改动 $p(x)$ 不会影响 (1.19) 式中积分的值. 所以连续型分布函数的密度不是唯一确定的. 但是反过来说, 满足 (1.20) 式中 $1^\circ, 2^\circ$ 两条性质的密度 $p(x)$ 却可通过 (1.19) 式完全确定分布函数 $F(x)$, 从而完全确定 ξ 的概率分布 $\mathbf{F}$. 事实上, 此时有 $dF(x) = p(x)\, dx$. 因而 (1.15) 式就化为

$$P\{\xi \in B\} = \int_B p(x)\, dx, \qquad 任\ \ B \in \mathcal{B}. \tag{1.21}$$

连续型随机变量定义中的 (1.19) 式只不过是此式中 $B = (-\infty, x)$ 时的一个特殊形式. 此外尚有

$$P\{\xi \in [a,b)\} = \int_a^b p(x)\, dx, \qquad 任\ \ a < b, \tag{1.22}$$

上式左端的区间 $[a, b)$ 换为 (a, b),$(a, b]$ 或 $[a, b]$ 后仍真.

连续型分布的密度函数与离散型分布中的密度阵相对照，无论就其性质或所起的作用而言，都是地位相当的. 只是在离散型情形有 $P\{\xi = x_k\} = p_k$ ，而对连续型随机变量却永有 $P\{\xi = x\} = 0$. 但是，稍微改变叙述方式，我们仍可看到 $p(x)$ 的相应作用. 将 (1.22) 式用于区间 $[x, x + \Delta x]$ 并用定积分中值定理可知，当 Δx 很小时有

$$P\{\xi \in [x, x + \Delta x]\} = \int_x^{x+\Delta x} p(t)\, dt \approx p(x)\Delta x.$$

于是我们可以说，密度 $p(x)$ 的值反映了 ξ 落入 x 附近之概率的大小.

最后简要地说明，离散型与连续型只不过是随机变量的两种特殊类型，而不是随机变量的分类. 分析中如下著名定理表明，还有许多其它类型的随机变量.

定理 1.7 勒贝格分解 分布函数 $F(x)$ 有分解

$$F(x) = c_1F_1(x) + c_2F_2(x) + c_3F_3(x), \tag{1.23}$$

其中常数 $c_1, c_2, c_3 \geq 0$, $c_1 + c_2 + c_3 = 1$, 而 $F_1(x), F_2(x), F_3(x)$ 都是分布函数， $F_1(x)$ 为纯跳跃函数， $F_2(x)$ 是绝对连续函数， $F_3(x)$ 为奇异函数.

定理 1.7 中奇异函数的含义及定理的证明可参见一般的实变函数论教科书，这里不能详述. 我们仅指出，分解式 (1.23) 中取 $c_1 = 1, c_2 = c_3 = 0$ 或取 $c_2 = 1, c_1 = c_3 = 0$ 就得到我们讨论过的离散型与连续型分布函数. 此外尚有奇异型及各种不同的混合型. 但最重要的还是我们讨论过的两种类型，我们将在下面进一步介绍这两类分布的常见例子.

2.1.5 习题

1. 设随机变量 ξ 的分布函数为 $F(x)$, 证明 $\eta = e^{\xi}$ 也是随机变量，并求 η 的分布函数.

2. 设 A, B 为两个不相容随机事件，a, b 为两个常数，$a < b$，证明 $\xi = a\chi_A + b\chi_B$ 为随机变量，并求 ξ 的分布函数.

3. 若对每个 $n \geq 1$，ξ_n 为随机变量，证明：

$$\sup_n \xi_n, \quad \inf_n \xi_n, \quad \overline{\lim_n} \xi_n, \quad \underline{\lim_n} \xi_n$$

均为随机变量.

4. 假定一硬币抛出正面的概率为 $p(0 < p < 1)$，反复抛这枚硬币直至正面与反面都出现过为止，求 (1) 抛掷次数恰为 k 的概率 p_k; (2) 恰好抛偶数次的概率.

5. 在半径为 R 的圆内任取一点 (二维几何模型)，求此点到圆心之距离 ξ 的分布函数及 $P\{\xi > 2R/3\}$.

6. 已知随机变量 ξ 的密度函数为

$$p(x) = \begin{cases} x, & 0 < x \leq 1, \\ 2 - x, & 1 < x \leq 2, \end{cases}$$

试求 (1) ξ 的分布函数，(2) $P\{0.2 < \xi < 1.2\}$.

7. 设

$$p(x) = e^{-e(x-a)}, \qquad x > 0.$$

(1) 求 a 使 $p(x)$ 为密度函数；(2) 若 ξ 以此 $p(x)$ 为密度，求 b 使 $P\{\xi > b\} = b$.

8. 设 $F_1(x)$ 与 $F_2(x)$ 为两个分布函数，c_1 与 c_2 为二常数，$c_1, c_2 \geq 0$，$c_1 + c_2 = 1$. 求证 $c_1F_1(x) + c_2F_2(x)$ 也是分布函数.

9. 设 $F(x)$ 是连续型分布函数，试证对任意 $a < b$ 有

$$\int_{-\infty}^{+\infty} [F(x+b) - F(x+a)]\, dx = b - a.$$

10. 设

$$F(x) = \begin{cases} (1+2x)/3, & 0 < x \leq 1 \\ 1, & 1 < x. \end{cases}$$

试证明 (1)$F(x)$ 是一个分布函数； (2)$F(x)$ 既不是离散型也不是连续型的，但它可以写为这两种类型分布函数的线性组合.

2.2 贝努里概型及其中的离散型分布

2.2.1 贝努里试验

只有两种可能结果的试验称作**贝努里试验**. 例如掷两枚骰子，可能得到"同点"也可能得到"不同点"；抽取一件产品检验，可能是"正品"也可能是"次品"等等都是贝努里试验. 注意，这种试验的样本空间 Ω 不一定只含两个样本点，只是我们把所关心的一部分样本点归结为一种结果 A，同时把其余样本点的集合看作另一种结果 $\overline{A}$. 在上述掷两枚骰子的例子中，样本空间 Ω 含 36 个样本点，我们只关心"同点"是否发生，于是把其中 6 个样本点组成的事件 $A=\{(i,i):i=1,2,\cdots,6\}$ 视为一种结果，其余 30 个样本点组成另一结果 $\overline{A}=\{$ 不同点 $\}$，此外，我们不再关心由 Ω 的非空真子集组成的其它事件. 于是，对于贝努里试验而言，事件 σ 代数应取为 $\mathcal{F}=\{\varnothing, A, \overline{A}, \Omega\}$. 通常把结果 A 称作"成功"，而把 $\overline{A}$ 称作"失败". 再取定成功与失败的概率

$$p=P(A) \text{ 与 } q=P(\overline{A}), \quad p,q>0 \quad \text{且 } p+q=1, \tag{2.1}$$

则建立了一次贝努里试验的概率空间 $(\Omega,\mathcal{F},P)$.

在概率论的理论及应用中，经常以一系列独立重复的贝努里试验作为概率模型. 独立重复试验的严格定义已于 1.6.4 节中叙述. 粗略地说，所谓重复，即各次试验的概率空间都是上述 $(\Omega,\mathcal{F},P)$. 而 n 个试验的独立性是指各次试验的结果互不影响，即对第 i 次试验的任何结果 $E_i\ (i=1,2,\cdots,n)$ 都有

$$P(E_1E_2\cdots E_n)=P(E_1)P(E_2)\cdots P(E_n). \tag{2.2}$$

这样的 n 个独立重复的贝努里试验称作 **n重贝努里试验**. 将一贝努里试验独立地重复下去所得的一系列试验，就是**可列重贝努里试验**.

2.2.2 二项分布

考虑 n 重贝努里试验中成功次数 ξ, 易见 ξ 的可能值为 $k = 0, 1, 2, \cdots, n$. 注意 $\{\xi = k\}$ 当且仅当这 n 次试验中恰有 k 个成功 A 与 $n-k$ 个失败 $\overline{A}$. 先考虑前 k 次试验全成功而后 $n-k$ 次试验全失败这一特殊情形. 由 (2.1) 与 (2.2) 式可得出现这种结果的概率

$$P\{\underbrace{A\cdots A}_{k\text{个}}\underbrace{\overline{A}\cdots\overline{A}}_{n-k\text{个}}\} = P(A)\cdots P(A)P(\overline{A})\cdots P(\overline{A}) = p^k q^{n-k}.$$

注意所得结果仅与 A 的个数 k 有关, 与 A 出现在哪 k 个位置上无关. 再者, 在这 n 次试验中选择 k 次成功共有 $\binom{n}{k}$ 种方式, 且各种方式两两不相容, 故由可加性立得 ξ 的密度

$$P\{\xi = k\} = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \cdots, n.$$

一般地, 任给定自然数 n 及正数 $p, q\,(p+q=1)$, 令

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \cdots, n. \tag{2.3}$$

则 $b(k; n, p) > 0$ 且

$$\sum_{k=0}^{n} b(k; n, p) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1.$$

称以 $\{b(k; n, p)\}$ 为密度的离散型分布为**二项分布**. 当 $n = 1$ 时的特例又称作**贝努里分布**. 这是一个两点分布, 其密度阵为 $\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$.

上述推导表明, n 重贝努里试验的成功次数 ξ 服从参数为 n, p 的二项分布.

下面讨论二项分布的性质. 对 $k \geq 1$, 考虑比值

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p - k}{kq}.$$

易见，当 $k < (n+1)p$ 时，$b(k;n,p) > b(k-1;n,p)$，而当 $k > (n+1)p$ 时，$b(k;n,p) < b(k-1;n,p)$. 这说明，对任何固定的参数 n 与 p, $b(k;n,p)$ 的值先随 k 的变大而上升，再随 k 的变大而下降. 于是必有最大值. 如果 $m=(n+1)p$ 是整数，则 $b(m;n,p)=b(m-1;n,p)$ 同为 $b(k;n,p)$ 的最大值. 如果 $(n+1)p$ 不是整数，则 $b(k;n,p)$ 在 $m=[(n+1)p]$ 处取到最大值 (这里 $[a]$ 表示不超过 a 的整数). 图 2.1 给出二项分布这一性质的图示. 我们称使 $b(k;n,p)$ 取到最大值的 m 为二项分布随机变量的**最可能值**，或称为 n 重贝努里试验的**最可能成功次数**.

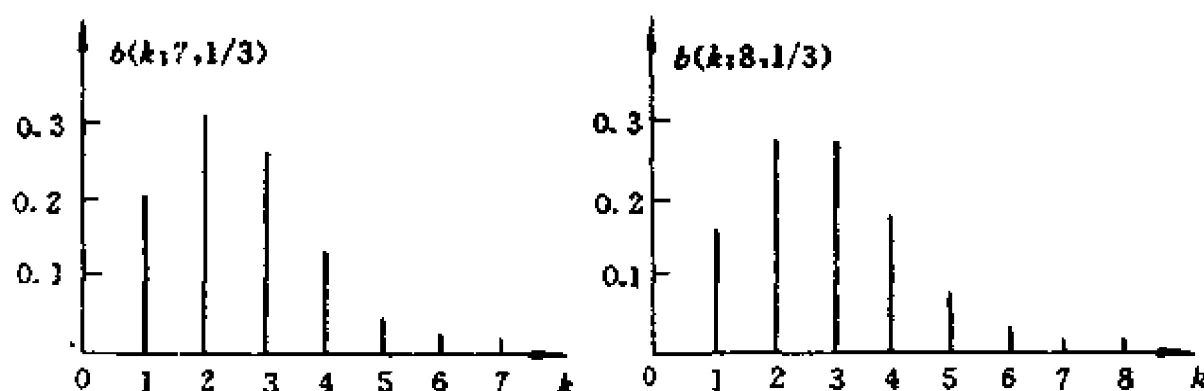


图 2.1 二项分布的最可能值

例 2.1 抛一均匀硬币 $2n$ 次是 $p=1/2$ 的 $2n$ 重贝努里试验. 因为 $(2n+1)p=n+1/2$ 不是整数，故最可能成功次数为 $[(2n+1)p]=n$. 就是说，$2n$ 次抛掷中正反面各出现 n 次的可能性最大. 利用斯特灵近似公式

$$n! \approx n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$$

可得此二项概率的最大值为

$$b\left(n;2n,\frac{1}{2}\right)=\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\approx\frac{1}{\sqrt{n\pi}}.$$

因此，当试验次数 n 无限变大时，二项分布的最可能值 (从而每个可能值) 出现的概率趋于 0.

例 2.2 小概率事件 在可列重贝努里试验中，求事件 $E=\{$ 试验终将成功 $\}$ 的概率.

解: 令 $F_n = \{$ 前 n 次试验均失败 $\}$, 则 $\{F_n\}$ 为单调下降事件列, 且

$$\lim_n F_n = \bigcap_n^{\infty} F_n = \overline{E}.$$

利用概率的上连续性及 $P(F_n) = b(0; n, p) = q^n$ 可得

$$P(E) = 1 - P(\lim_n F_n) = 1 - \lim_n P(F_n) = 1 - \lim_n q^n.$$

于是, 不论成功的概率 p 多么小, 总有 $0 < q < 1$, 从而 $P(E) = 1$. 就是说小概率事件终将发生的概率是 1.

下述例题给出二项分布的一个实际应用.

例 2.3 设每台自动机床在运行过程中需要维修的概率均为 $p = 0.01$, 并且各机床是否需要维修相互独立. 如果 (1) 每名维修工人负责看管 20 台机床; (2) 3 名维修工人共同看管 80 台机床, 求不能及时维修的概率.

解: (1) 这是 $n = 20$ 重贝努里试验, 参数 $p = 0.01$, 需维修的机床数 ξ 服从 $b(k; 20, 0.01)$ 分布. 故不能及时维修的概率为

$$P\{\xi > 1\} = 1 - b(0; 20, 0.01) - b(1; 20, 0.01)$$

$$= 1 - 0.99^{20} - 20 \times 0.01 \times 0.99^{19} \approx 0.0169.$$

(2) 类似可得 3 名维修工人共同看管 80 台机床时不能及时维修的概率

$$P\{\xi > 3\} = 1 - \sum_{k=0}^{3} b(k; 80, 0.01) \approx 0.0087.$$

比较本例的 (1) 与 (2) 可见, 3 名维修工人共同看管 80 台机床, 比每人看管 20 台机床提高了工作效率. 但是, 不能及时维修的概率却下降到将近一半. 因而, 从这一点来看还是 “联合起来好”. 象这种不太复杂的计算有时也会给经营管理者带来好的效益. 但是必须指出, 在解这个例题中, 我们忽略了 “维修时间” 这个因素. 所得结

果只适用于服务时间很短的服务系统. 例如, 纺纱厂的接头工人, 人工交换台的接线员, 小商品的售货员等等. 如果再考虑常值或随机长度的服务时间, 问题会复杂得多. 有关问题的研究属于概率论的一个卓有成效的分支排队论.

易见, n 越大有关二项分布的计算量越大. 比如在上例中, 如果某车间有这种机床 200 台, 为使不能及时维修的概率在 0.02 以下, 问至少需多少名维修工人? 由于需维修的机床数 ξ 服从 $b(k;200,0.01)$ 分布. 问题化为求维修工人数 r, 使

$$P\{\xi > r\} = \sum_{k=r+1}^{200} b(k;200,0.01) \le 0.02.$$

直接计算是很麻烦的, 我们将在 2.3.1 节给出它的近似算法.

2.2.3 几何分布

考虑可列重贝努里试验中首次成功的等待时间 ξ. 它取值自然数 k, 当且仅当前 $k-1$ 次试验全失败同时第 k 次成功. 于是由 (2.1) 及 (2.2) 式可得 ξ 的密度

$$P\{\xi = k\} = P\{\overline{A}\cdots\overline{A}A\} = q^{k-1}p, \qquad k = 1,2,\cdots.$$

一般地, 任给参数 p $(0 < p < 1,\ q = 1-p)$ 令

$$g(k;p) = q^{k-1}p, \qquad k = 1,2,\cdots. \tag{2.4}$$

那么有 $g(k;p) > 0$ 且 $\sum\limits_{k=1}^{\infty} g(k;p) = \sum\limits_{k=1}^{\infty} q^{k-1}p = p/(1-q) = 1$. 于是 $\{g(k;p)\}$ 可以作为离散型分布的密度, 称它为**几何分布**.

几何分布有如下本征性质:

定理 2.1 取值于自然数的随机变量 ξ 有几何分布, 当且仅当 ξ 有无记忆性:

$$P\{\xi > m+n \mid \xi > m\} = P\{\xi > n\}, \qquad 任\quad m,n \ge 1. \tag{2.5}$$

证明: 如果 ξ 的密度如 (2.4) 式, 则

$$P\{\xi > n\} = \sum_{k=n+1}^{\infty} q^{k-1}p = \frac{pq^n}{1-q} = q^n, \qquad 任 \quad n \geq 1.$$

于是对任意的 $m, n \geq 1$ 有

$$P\{\xi > m+n \,|\, \xi > m\} = \frac{P\{\xi > m+n\}}{P\{\xi > m\}} = \frac{q^{m+n}}{q^m} = P\{\xi > n\}.$$

即无记忆性 (2.5) 式成立.

另一方面, 由 (2.5) 式知

$$Q_n = P\{\xi > n\} > 0, \qquad 任 \quad n \geq 1.$$

并且有

$$Q_{m+n} = Q_m Q_n, \qquad 任 \quad m, n \geq 1.$$

解这个方程可得 $Q_m = Q_1^m$. 注意, 一方面 $Q_1 > 0$, 另一方面 $Q_1 = 1$ 将导致一切 $Q_m = 1$, 这与 ξ 取值自然数矛盾, 故取 $q = Q_1$, 满足 $0 < q < 1$. 且对任意 $k \geq 1$ 有

$$P\{\xi = k\} = Q_{k-1} - Q_k = q^{k-1} - q^k = q^{k-1}p.$$

得证 ξ 服从几何分布. 定理证完.

设 ξ 为可列重贝努里试验中首次成功的等待时间. 上述无记忆性 (2.5) 表明: 已知试验了 m 次未获成功, 再加做 n 次试验仍不成功的条件概率, 等于从开始算起 n 次试验都不成功的概率. 就是说, 已做过的 m 次失败的试验被忘记了. 产生几何分布这种无记忆性的根本原因在于, 我们进行的是独立重复试验, 这是不学习, 不总结经验的一系列试验, 当然已做过的试验不会留下记忆. 常言说 "失败是成功之母", 其前提条件就是每次试验后要认真总结经验, 不断改进试验方案, 才能尽快地取得成功. 如果真的在做 "独立重复试验", 那么不管已经失败过多少次, 也不会为今后的试验留下可借鉴的东西.

例 2.4 十把外形相同的钥匙中只有一把能打开门. 现任意一一试开, 试对每次试毕 (1) 放回; (2) 不放回 二情形求事 E = {至多试 3 次打开门} 的概率.

解: (1) 放回情形是独立重复试验, 属贝努里概型. 以 ξ 表首次打开门的等待时间, 它服从 $p = 0.1$ 的几何分布. 于是所求概率

$$P(E) = P\{\xi \leq 3\} = g(1;0.1) + g(2;0.1) + g(3;0.1) = 0.271.$$

(2) 不放回情形不再是独立重复试验, 适用古典概型. 样本点总数 $n(\Omega) = \binom{10}{3}$, 而 $n(\overline{E}) = \binom{9}{3}$. 故

$$P(E) = 1 - P(\overline{E}) = 1 - \frac{\binom{9}{3}}{\binom{10}{3}} = 0.3.$$

这个简单的例子再次说明, 正确地判断面临的具体问题适用何种概型, 是解决问题的前提.

2.2.4 巴斯卡分布

考虑可列重贝努里试验中第 r 次 $(r \geq 1)$ 成功的等待时间 ξ_r. 易见 ξ_r 的可能值是 $k = r, r+1, \cdots$, 并且有

$$\{\xi_r = k\} = \{\text{前 } k-1 \text{次试验恰有 } r-1 \text{次成功且第 } k \text{次成功}\}.$$

由 (2.1) 与 (2.2) 式可得

$$\begin{aligned} P\{\xi = k\} &= P\{\text{前}k-1\text{次中恰成功}r-1\text{次}\} \cdot P\{\text{第}k\text{次成功}\} \\ &= b(r-1;k-1,p) \cdot p = \binom{k-1}{r-1} p^r q^{k-r}. \end{aligned}$$

一般地, 对 $r \geq 1$, $p, q > 0, p + q = 1$, 令

$$f(k;r,p) = \binom{k-1}{r-1} p^r q^{k-r}, \qquad k = r, r+1, \cdots, \tag{2.6}$$

则有 $f(k;r,p)>0$ 且

$$\sum_{k=r}^{\infty} f(k;r,p)=\sum_{k=r}^{\infty}\binom{k-1}{r-1}p^r q^{k-r}$$

$$=p^r\sum_{i=0}^{\infty}\binom{r+i-1}{r-1}q^i=p^r(1-q)^{-r}=1.$$

可见数列 $\{f(k;r,p)\}$ 可以成为是离散型分布的密度. 称作参数为 r,p 的**巴斯卡分布**, 或称为**负二项分布**.

由 (2.4) 式定义的几何分布是 $r=1$ 时巴斯卡分布的特例, 即 $g(k;p)=f(k;1,p)$. 另一方面, 如记 $\tau_1=\xi_1,\tau_n=\xi_n-\xi_{n-1}(n>1)$. 则随机变量 τ_n 是第 $n-1$ 次成功到第 n 次成功的间隔时间. 显然有

$$\xi_r=\tau_1+\tau_2+\cdots+\tau_r. \tag{2.7}$$

以后我们将看到, $\tau_1,\tau_2,\cdots,\tau_r$ 是 r 个相互独立的随机变量, 而且每个 τ_k 均服从几何分布.

下面是应用巴斯卡分布的两个著名的例题.

例 2.5 巴拿赫火柴问题 某人口袋中有两盒火柴, 开始时每盒各装 n 根. 每次他从口袋中任取一盒使用其中一根火柴. 求此人掏出一盒发现已空, 而另一盒尚余 r 根的概率.

解: 记 $E=\{$ 掏出甲盒已空而乙盒尚余 r 根 $\}$. 由对称性可知所求概率为 $2P(E)$. 我们以取出甲盒为 "成功". 这是一个 $p=1/2$ 的独立重复贝努里试验. 而 $E=\{$ 第 $n+1$ 次成功发生在第 $2n-r+1$ 次试验 $\}$. 故所求概率为

$$2P(E)=2f(2n-r+1;n+1,\frac{1}{2})=\binom{2n-r}{n}2^{r-2n}.$$

例 2.6 分赌注问题 在可列重贝努里试验中, 求事件 $E=\{n$ 次成功发生在 m 次失败之前 $\}$ 的概率.

解: 记 $F_k = \{$ 第 n 次成功发生在第 k 次试验 $\}$, 则有

$$E = \bigcup_{k=n}^{n+m-1} F_k,$$

且右方诸 F_k 是两两不相容的, 故所求概率

$$P(E) = \sum_{k=n}^{n+m-1} P(F_k) = \sum_{k=n}^{n+m-1} f(k; n, p) = \sum_{k=n}^{n+m-1} \binom{k-1}{n-1} p^n q^{k-n}. \tag{2.8}$$

这个例子起源于概率论历史上著名的"分赌注问题".1654 年, 当时的职业赌徒德梅雷爵士向法国大数学家巴斯卡提出如下问题: 甲乙二人各下赌注 d 元, 商定先胜三局者赢得全部赌金. 假定在每一局中二人获胜机会相等, 且各局胜负相互独立. 如果当甲胜一局而乙尚未获胜时赌博被迫中止, 问赌注应当怎么分?

为解决这个问题, 巴斯卡与当时声望很高的数学家费尔马建立了通信联系. 他们的卓有成效的讨论, 不仅完满回答了分赌注问题, 而且为解决其它有关概率问题搭起了框架. 同时, 他们二人的联系, 吸引了不少数学家参加相关的讨论, 激发了欧洲的数学家对概率论的兴趣, 从而极大地促进了概率论的建立与发展. 正因为如此, 这个例题在概率论历史上占有荣誉的地位.

巴斯卡令人信服地指出, 赌金的分法应当取决于若赌博能持续下去甲与乙各自获胜的概率. 易见, 甲将赢得全部赌注的概率, 就是在 $p = 0.5$ 的可列重贝努里试验中, 2 次成功在 3 次失败之前出现的概率. 用 (2.8) 式的结果可算得此概率为 11/16. 而乙将赢得全部赌注的概率为 5/16. 因此, 总额为 $2d$ 的赌注应按 11 : 5 的比例分给甲乙二人.

2.2.5 习题

1. 向目标进行 20 次独立的射击, 假定每次命中率均为 0.2. 试求 (1) 至少命中 1 次的概率; (2) 至多命中 2 次的概率; (3) 最可能命中次数.

2. 同时掷两枚骰子，直到某个骰子出现 6 点为止．求恰好掷 n 次的概率．

3. 某公司经理拟将一提案交董事代表会批准，规定如提案获多数代表赞成则通过．经理估计各代表对此提案投赞成票的概率为 0.6，且各代表投票情况相互独立．为以较大概率通过提案，试问经理请 3 名董事代表好还是请 5 名好？

4. 求 n 重贝努里试验中成功偶数次的概率 P_n．

5. 甲，乙二队比赛篮球．假定每一场甲乙队获胜的概率分别为 0.6 与 0.4，且各场胜负独立．如果规定先胜 4 场者为冠军，求甲队经 i 场 $(i = 4,5,6,7)$ 比赛而成为冠军的概率 p_i．再问与"三场两胜"制相比较，采用哪种赛制甲队最终夺得冠军的概率较小？

6. 在巴拿赫火柴问题中，求当从某盒中取出最后一根时，另一盒尚余 r 根的概率．

7. 在可列重贝努里试验中，以 ξ_i 表第 i 次成功的等待时间，求证 $\xi_2 - \xi_1$ 与 ξ_1 有相同的概率分布．

8. **广义贝努里试验** 假定一试验有 r 个可能结果 $A_1,\cdots,A_r$，并且

$$P(A_i) = p_i > 0, \quad p_1 + \cdots, p_n = 1.$$

现将此试验独立地重复 n 次．求 A_1 恰出现 k_1 次，$\cdots,A_r$ 恰出现 k_r 次 $(k_i \geq 0, k_1 + \cdots + k_r = n)$ 的概率．

9. 续上题．在可列重广义贝努里试验中，求 A_i 在 A_j 之前出现的概率 $(i \neq j)$．

2.3 普阿松分布

2.3.1 普阿松定理

普阿松分布是概率论中一种重要的离散型分布，它在理论和实践中都有广泛的应用．下面将从两个不同的角度引进这一分布，并初步讨论它的性质．先从二项分布的近似计算说起．

在上节例 2.3 中我们已经看到，当参数 n 较大时，有关二项分布 $b(k;n,p)$ 的计算较为麻烦．普阿松给出的如下定理提供了计算二项概率的一种近似方法．

定理 3.1 设有一列二项分布 $\{b(k;n,p_n)\}$，其参数列 $\{p_n\}$ 满足

$$\lim_n np_n = \lambda > 0, \tag{3.1}$$

则对任何非负整数 k 有

$$\lim_n b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}. \tag{3.2}$$

证明: 直接计算可得

$$b(k;n,p_n) = \frac{n!}{k!(n-k)!}\, p_n^k\, (1-p_n)^{n-k}$$

$$= \frac{1}{k!}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})(1-p_n)^{-k}(np_n)^k(1-p_n)^n.$$

固定 $k \geq 0$ 令 $n \to \infty$，上式右方除最后两个因子外，其余各含 n 的因子均趋于 1，而 $(np_n)^k \to \lambda^k$，余下的只要再证

$$\lim_n (1-p_n)^n = e^{-\lambda},$$

这又只需证明序列 $\{(1-p_n)^n\}$ 与 $\{(1-\lambda/n)^n\}$ 有共同的极限．因为 n 充分大时有 $|\,1-p_n\,| \leq 1$ 及 $|\,1-\lambda/n\,| \leq 1$，运用 $|a| \leq 1$ 且 $|b| \leq 1$ 时的初等不等式 $|\,a^n - b^n\,| \leq n\,|\,a-b\,|$，并利用 (3.1) 式便得到

$$|\,(1-p_n)^n - (1-\lambda/n)^n\,| \leq n|\,(1-p_n)-(1-\lambda/n)\,|$$

$$= |\,np_n - \lambda\,| \to 0.$$

至此定理得证．

有了普阿松定理，当 n 很大而且 p 很小时，可以用近似公式

$$b(k;n,p) \approx \frac{(np)^k}{k!}e^{-np}. \tag{3.3}$$

当然，n 越大误差越小．同时，为了适应 (3.1) 式的需要，我们要求 p 很小，以保持乘积 np 有适度的大小．

对参数 $\lambda > 0$, 记

$$p(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k = 0, 1, 2, \cdots. \tag{3.4}$$

易见 $p(k;\lambda) > 0$ 且 $\sum_{k=0}^{\infty} p(k;\lambda) = \sum_{k=0}^{\infty} \lambda^k/k! \cdot e^{-\lambda} = e^{\lambda}e^{-\lambda} = 1$. 即数列 $\{p(k;\lambda)\}$ 可成为离散型分布的密度，将它称作以 λ 为参数的**普阿松分布**. $p(k;\lambda)$ 的数值有现成的表可以查到 (参见本书末附表 II).

回到上节例 2.3 的 (3)．全车间需维修的机床数 ξ 服从 $b(k;\, 200, 0.01)$ 分布．运用 (3.3) 式，可按 ξ 服从 $\lambda = 200 \times 0.01 = 2$ 的普阿松分布计算．查表可得 $r = 5$ 时不能及时维修的概率

$$P\{\xi > 5\} = \sum_{k=6}^{\infty} p(k;2) = 0.0166,$$

就是说，安排 5 名维修工人，便可以使不能及时维修的概率在 0.02 之下．若用二项分布直接计算，则有

$$P\{\xi > 5\} = 1 - \sum_{k=0}^{5} b(k;200,0.01) = 0.0160.$$

所得结果相差甚微，结论仍然是需要 5 名维修工人便可达到要求．

在应用概率论中，诸如服务系统中对服务的呼唤数，产品的缺陷 (如布匹上的疵点) 数，一定时期内出现的稀有事件 (如意外事故，灾害等) 个数，放射性物质发射出的粒子数等等都以普阿松分布为其概率模型．这是因为，上述例子本来就是 n 大 p 小的二项分布．以服务系统中的呼唤数为例．服务设施的用户 n 很大，每个用户在指定时间内使用这个设施的概率 p 甚小，而且各用户使用情况又独立．于是，对此服务设施的呼唤数应是 n 大 p 小的二项分布．用普阿松定理，可近似认为服从 $\lambda = np$ 的普阿松分布．上述应用例子表明，普阿松分布广泛存在于社会生活的许多方面，它在运筹学及管理科学中占有突出的地位．

关于放射粒子数服从普阿松分布，卢瑟福的著名的实验提供了很好的证明．近代物理学家卢瑟福等人利用他们设计的云雾室对放射性物质镭所发射出的 α 粒子进行计数．他们记录下每个长 7.5 秒的时段内观测到的 α 粒子的个数，共计观测了 $n=2608$ 个时段，其中记录下 k 个粒子的时段的频数为 ν_k．另一方面，这 2608 个时段共记录到 10097 个 α 粒子，平均每个长 7.5 秒的时段记录 3.87 个粒子．以此作为 λ 的估计值（这样做的理由将在下章说明）．就是说，认为每个时段放射 α 粒子数服从 $p(k;3.87)$ 分布．据此，对不同的 k 算出记录 k 个粒子的时段的理论频数 $n\cdot p(k;3.87)$．表 3.1 对照地列出频数的实验观测值与理论值，可见它们非常接近．这就说明以普阿松分布作为放射粒子数的概率模型是合适的．

表 3.1　卢瑟福实验值与理论值对照

k	ν_k	$n\cdot p(k;3.87)$
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
≥ 10	16	17.076
总　计	2608	2608.001

2.3.2　普阿松分布的性质

普阿松分布的密度值 $p(k;\lambda)$ 随 k 的变化情况与二项分布类似．考虑比值

$$\frac{p(k;\lambda)}{p(k-1;\lambda)} = \frac{\lambda^k (k-1)! e^{-\lambda}}{\lambda^{k-1} k! e^{-\lambda}} = \frac{\lambda}{k}, \qquad k \geq 1. \tag{3.5}$$

当 $k < \lambda$ 时有 $p(k-1;\lambda) < p(k;\lambda)$, 而当 $k > \lambda$ 时 $p(k-1;\lambda) > p(k;\lambda)$. 因此, 随着 k 由小变大, $p(k;\lambda)$ 的值先上升后下降. $p(k;\lambda)$ 在 $m = [\lambda]$ 处达最大值, 当 λ 为整数时, $p(k;\lambda)$ 在 $m = \lambda$ 及 $m-1$ 处达到其最大值. 这个 m 称作普阿松分布的**最可能值**.

我们指出, (3.5) 式也是普阿松分布的充分条件, 其证明留作习题. 下面结合放射粒子数的例, 介绍普阿松分布的一个重要性质.

例 3.1 普阿松分布在随机选择下不变 假设一块放射性物质在单位时间内发射出的 α 粒子数 ξ 服从参数为 λ 的普阿松分布. 而每个放射出的 α 粒子被记录下来的概率均为 p, 就是说有 $(1-p)$ 的概率被计数器遗漏. 如果各粒子是否被记录相互独立, 试求记录下的 α 粒子数 η 的分布.

解: 以事件 $\{\xi = n\}, n = 0, 1, 2, \cdots$ 为分割用全概公式可得, 对非负整数 k 有

$$P\{\eta = k\} = \sum_{n=0}^{\infty} P\{\xi = n\} P\{\eta = k \mid \xi = n\}$$

$$= \sum_{n=k}^{\infty} p(n;\lambda)\, b(k;n,p) = \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{n=k}^{\infty} \frac{(\lambda q)^{n-k}}{(n-k)!} \frac{1}{k!} e^{-\lambda} (\lambda p)^k = \frac{1}{k!} (\lambda p)^k e^{-\lambda p}.$$

于是, 记录下的 α 粒子数 η 仍服从普阿松分布, 只是其参数为 λp, 比 ξ 的参数相应地多了一个折扣因子 p. 一般的普阿松分布都具备这种随机选择不变性.

2.3.3 普阿松过程

本节介绍普阿松分布的机制, 从而从另一角度引进普阿松分布.

假定有一个于随机时刻陆续到来的质点流. 这里的“质点”可以是前面例子中提到的呼唤, 意外事故或 α 粒子等. “于随机时间陆续到达”是指质点一个一个地到达, 但质点到达的时间间隔都是随机变量. 对任何 $t \geq 0$, 以 ξ_t 代表在 $[0,t)$ 时段内到达的质点个数. 每个 ξ_t 都是非负整值随机变量. 下面我们将证明, 添加一组通常的假设, $\{\xi_t, t \geq 0\}$ 是所谓的**普阿松过程**.

定理 3.2 假定于随机时间陆续到达的质点流满足以下各条件:

1°**独立增量性** 在不相交时段内到达的质点数目相互独立. 即对任何 $0 \leq t_1 < t_2 \leq t_3 < t_4$, 任何 $k_1, k_2 \geq 0$, 事件 $\{\xi_{t_2} - \xi_{t_1} = k_1\}$ 与 $\{\xi_{t_4} - \xi_{t_3} = k_2\}$ 相互独立.

2°**平稳性** 在长为 t 的时段 $[a, a+t)$ 内到达 k 个质点的概率, 只与计时长度 t 有关而与计时起点 a 无关. 于是可记

$$P_k(t) = P\{\xi_{a+t} - \xi_t = k\}. \tag{3.6}$$

3°**普通性** 在有限的时间区间内只来有限个质点, 即对任 $t > 0$ 有 $\sum\limits_{k=0}^{\infty} P_k(t) = 1$. 并且在充分短的时间 t 内最多只来一个质点, 即假设多于一个质点到达的概率 $\sum\limits_{k=2}^{\infty} P_k(t) = o(t)$. 并排除总也不来质点这种无意义的情形, 即假定 $P_0(t)$ 不恒等于 1.

那么, 必存在常数 $\lambda > 0$, 使对一切 $t > 0$ 有

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad k = 0, 1, 2, \cdots. \tag{3.7}$$

证明: 对任何 $t, \Delta t > 0$ 考察 $P_k(t + \Delta t)$. 运用全概率公式及独立增量性 1° 可得

$$P_k(t + \Delta t) = \sum_{i=0}^{k} P_{k-i}(t) P_i(\Delta t), \qquad k \geq 0. \tag{3.8}$$

特别 $k = 0$ 时, 上式化为

$$P_0(t + \Delta t) = P_0(t) P_0(\Delta t). \tag{3.9}$$

由于 $P_0(t)$ 是 $[0,t)$ 时段内无质点到达的概率，它关于 t 单调非增．故作为函数方程 (3.9) 的有界单调解 (参见本节习题 8)，存在常数 $a, 0 \le a \le 1$ 使

$$P_0(t) = a^t, \quad t \ge 0.$$

但是 $a=0$ 或 1 导致 $P_0(t)$ 恒为 0 或为 1，这与 3° 矛盾．故存在常数 $\lambda > 0$ 使 $a = e^{-\lambda}$，从而有

$$P_0(t) = e^{-\lambda t}, \qquad t \ge 0. \tag{3.10}$$

这正是 $k=0$ 时的 (3.7) 式．

以下用归纳法，设 $(k-1)$ 时 (3.7) 式为真，往证 k 时 (3.7) 式成立．当 Δt 很小时，用普通性 3° 知

$$0 \le \sum_{i=2}^{k} P_{k-i}(t) P_i(\Delta t) \le \sum_{i=2}^{\infty} P_i(\Delta t) = o(\Delta t),$$

于是 $P_k(t+\Delta t)$ 的表达式 (3.8) 可写为

$$P_k(t+\Delta t) = P_k(t)P_0(\Delta t) + P_{k-1}(t)P_1(\Delta t) + o(\Delta t). \tag{3.11}$$

但由已证明的 (3.10) 式有

$$P_0(\Delta t) = 1 - \lambda \Delta t + o(\Delta t).$$

从而再用普通性 3° 知

$$P_1(\Delta t) = 1 - P_0(\Delta t) - \sum_{i=2}^{\infty} P_i(\Delta t) = \lambda \Delta t + o(\Delta t).$$

将上述二式代入 (3.11) 式便得

$$\frac{1}{\Delta t}[P_k(t+\Delta t) - P_k(t)] + \lambda P_k(t) = \lambda P_{k-1}(t) + o(1),$$

令 $\Delta t \to 0$，并据归纳法假设将 $(k-1)$ 时的 (3.7) 式代入，就得到 $P_k(t)$ 满足的一阶微分方程

$$P_k'(t) + \lambda P_k(t) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t}. \tag{3.12}$$

联合明显的初值条件 $P_k(0)=0$ (注意这里 $k\geq 1$), 便得方程 (3.12) 的解为

$$P_k(t)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}.$$

定理于是得证.

满足此定理诸条件的质点流的计数过程 $\{\xi_t, t\geq 0\}$ 称作强度为 λ 的**普阿松过程**. 由 (3.6) 及 (3.7) 式可知, 任 $t>0$ 随机变量 ξ_t 服从普阿松分布

$$P\{\xi_t=k\}=\frac{(\lambda t)^k}{k!}e^{-\lambda t}, \qquad k=0,1,2,\cdots. \tag{3.13}$$

对普阿松过程的进一步研究, 是随机过程论的重要内容.

2.3.4 习题

1. 假定一本 500 页的书总共有 500 个错字, 每个错字等可能地出现在每一页上. 试用普阿松近似计算指定一页上有至少 3 个错字的概率.

2. 假定螺丝钉的废品率 $p=0.015$. 问一盒应当装多少只才能保证每盒中正品在 100 只以上的概率不小于 80%?

3. 据以往的记录, 某商店每月出售的电视机台数服从参数 $\lambda=7$ 的普阿松分布. 问月初应库存多少台电视机, 才能以 0.999 的概率保证满足顾客对电视机的需求?

4. 保险公司的资料表明, 持某种人寿保险单的人在保险期内死亡的概率为 0.005. 现售出这种保险单 1200 份, 求保险公司至多赔付 10 份的概率.

5. 假定每小时进入某商店的顾客数服从 $\lambda=200$ 的普阿松分布, 而进来的顾客将购买商品的概率均为 0.05, 且各顾客是否购物相互独立. 求在 1 小时中至少有 6 位顾客在此商店中购物的概率.

6. 设人工交换台每小时收到的电话呼唤是强度 $\lambda=60$ 的普阿松过程. 求接线员离开 30 秒而未耽误工作的概率.

7. 通过一交叉路口的汽车流可看作一个普阿松过程. 如果 1 分钟内没有汽车通过的概率为 0.02, 求 2 分钟内有多于 1 辆汽车通过的概率.

8. 试证明, 函数方程

$$f(x+y) = f(x)f(y)$$

的单调 (或连续) 解必为指数函数.

9. 试证, 如果非负整值离散型分布的密度 $\{p_k, k=0,1,\cdots\}$ 满足条件

$$\frac{p_k}{p_{k-1}} = \frac{\lambda}{k}, \qquad k \geq 1.$$

其中常数 $\lambda > 0$, 则此分布是以 λ 为参数的普阿松分布.

2.4 重要的连续性分布

2.4.1 均匀分布

如前所述, 几何概型中的样本点是在其样本空间中均匀分布的. 本节先就一维情形阐述它的严格定义. 回忆本章开头引入随机变量概念的例 1.2, 某人一觉醒来的时刻 ξ 是取值于区间 $(0,60)$ 的随机变量. 此 ξ 的概率分布具有如下特点: ξ 取值于 $\Omega=(0,60)$ 的任何波雷尔子集 A 的概率, 只有 A 的勒贝格测度成正比, 而与 A 的形状, 位置无关. ξ 取值的这种等可能性, 表现在其密度函数 $p(x)$ 在区间 $(0,60)$ 上取常数值 $1/60$, 而在区间 $(0,60)$ 之外为 0.

定义 4.1 任给定参数 $a<b$, 函数

$$p(x) = \frac{1}{b-a}, \qquad \text{当 } a < x < b \tag{4.1}$$

满足 (1.20) 式中的两个条件. 称以 (4.1) 式中 $p(x)$ 为密度函数的连续型分布为区间 (a,b) 上的**均匀分布**, 记作 $U(a,b)$.

易见, $U(a,b)$ 分布的分布函数为

$$F(x) = \begin{cases} \frac{x-a}{b-a}, & a < x < b, \\ \\ 1, & x \geq b. \end{cases} \tag{4.2}$$

运用 (4.2) 式可知，如果 ξ 服从 $U(a,b)$ 分布，则对任何 $a \leq x < y \leq b$ 有

$$P\{x \leq \xi < y\} = F(y) - F(x) = \frac{y-x}{b-a}.$$

这表明 ξ 落入 (a,b) 的任何子区间 (x,y) 内的概率只与 $y-x$ 有关. 可以证明，这个性质刻画了均匀分布，就是说，具备这一性质的随机变量必有均匀分布. 其证明留作习题.

图 4.1 画出了均匀分布的密度与分布函数的图形.

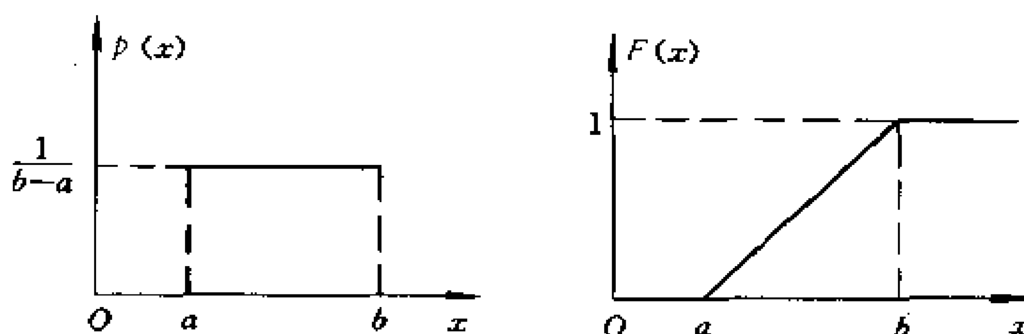


图 4.1 均匀分布的密度与分布函数

在应用中，定点计算的舍入误差被认为是均匀分布随机变量. 假定运算中的数据只保留小数点后 5 位，第 6 位四舍五入. 那么每次运算的舍入误差 ε 服从 $U(-0.5\times10^{-5}, 0.5\times10^{-5})$ 分布. 再者，假定班车每隔 a 分钟发出一辆，由于乘客不了解时刻表，他到达车站的时间是任意的 (具有等可能性)，故可以认为候车时间是区间 $(0,a)$ 上的均匀分布.

2.4.2 正态分布

对参数 a 与 $\sigma > 0$，令

$$\varphi_{a,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{4.3}$$

我们来证明这它是连续型的密度函数.

定理 4.1 由 (4.3) 式给出的 $\varphi_{a,\sigma}(x)$ 满足密度函数的条件 (1.20) 式.

证明: $\varphi_{a,\sigma}(x)>0$ 显然. 下计算定积分 $I=\int_{-\infty}^{+\infty}\varphi_{a,\sigma}(x)\,dx$. 注意 $\varphi_{a,\sigma}(x)$ 的原函数不是初等函数, 故不能用微积分基本公式计算 I. 现转而考察 I^2. 将累次积分化为二重积分并将积分变量替换为极坐标便可得到

$$I^2=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(y-a)^2}{2\sigma^2}}\,dy$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{2\pi}\,e^{-\frac{(s^2+t^2)}{2}}\,dsdt$$

$$=\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^{+\infty}e^{-r^2/2}r\,dr=1.$$

由此及 $I>0$ 得证 $I=1$. 定理证毕.

定义 4.2 以 (4.3) 式的 $\varphi_{a,\sigma}(x)$ 作密度函数的连续型分布称作参数为 a,σ^2 的**正态分布**, 记作 $N(a,\sigma^2)$. 特别地, 称 $N(0,1)$ 为**标准正态分布**, 并简写 $\varphi_{0,1}(x)$ 为 $\varphi(x)$. 正态分布又称为**高斯分布**.

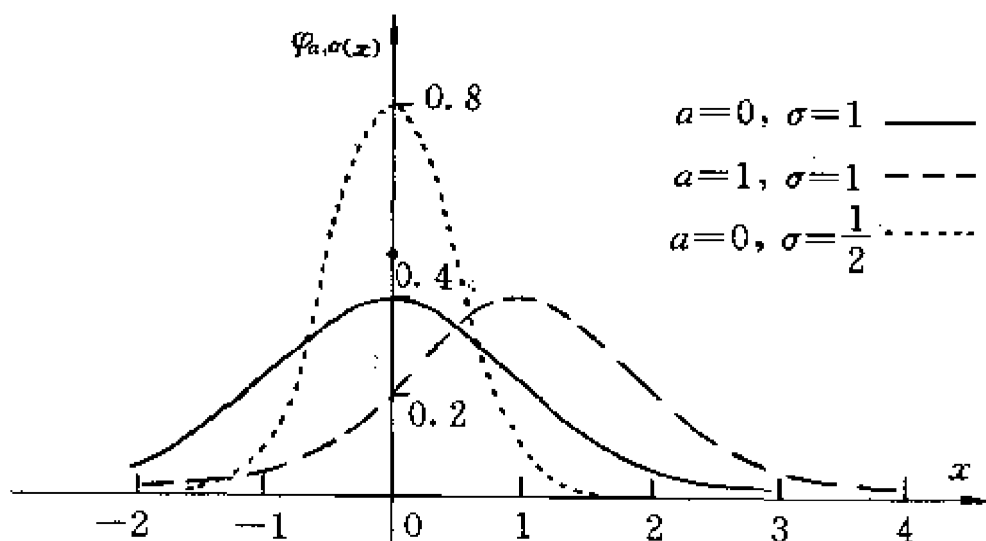


图 4.2 正态分布的密度曲线

图 4.2 画出了不同参数的正态分布密度函数的图形. 由图不难看到, 由 (4.3) 式给出的 $N(a,\sigma^2)$ 密度函数有如下性质:

1° 此分布关于参数 a 对称, 即有

$$\varphi_{a,\sigma}(a-x)=\varphi_{a,\sigma}(a+x), \qquad \text{任} \quad x\in\Re.$$

特别, $N(0,1)$ 密度 $\varphi(x)$ 为偶函数. 称 a 为此分布的**位置参数**.

2° $\varphi_{a,\sigma}(x)$ 在 $x=a$ 处取到最大值 $1/(\sigma\sqrt{2\pi})$. 注意到密度曲线下的面积应保持等于 1, 并且它在 x 处的值反映了此分布的取值 x 附近的概率大小. 于是, σ 越小, 密度的曲线越尖陡, 此分布取值越集中; σ 越大, 密度曲线越平缓, 此分布取值越分散. 通常称 σ 为正态分布的**形状参数**.

正态分布的分布函数为

$$\Phi_{a,\sigma}(x)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(t-a)^2}{2\sigma^2}}\,dt, \tag{4.4}$$

特别地, 标准正态分布的分布函数简记为 $\Phi(x)$. 一般正态分布的分布函数 $\Phi_{a,\sigma}$ 可以通过 $\Phi(x)$ 表出. 事实上将 (4.4) 式右方积分作变量替换 $u=(t-a)/\sigma$ 即可得到,

$$\Phi_{a,\sigma}(x)=\Phi(\frac{x-a}{\sigma}). \tag{4.5}$$

据此, 可将不同 x 对应的标准正态分布函数 $\Phi(x)$ 的值制成数值表 (见书末附表 III). 利用 (4.5) 式便可以得到各种参数正态分布函数 $\Phi_{a,\sigma}(x)$ 的值. 又据上述性质 1° 知 $\Phi(-x)=1-\Phi(x)$, 故只需对正数 x 列出相应的 $\Phi(x)$ 数值表就够用了.

例 4.1 3σ原则 设随机变量 ξ 服从 $N(a,\sigma^2)$ 分布, 求 $P\{|\xi-a|<3\sigma\}$.

解:

$$P\{|\xi-a|<3\sigma\}=P\{a-3\sigma<\xi<a+3\sigma\}$$

$$=\Phi_{a,\sigma}(a-3\sigma)-\Phi_{a,\sigma}(a-3\sigma)=\Phi(3)-\Phi(-3)=2\Phi(3)-1,$$

查表得 $\Phi(3)=0.9987$, 代入即得所求概率

$$P\{|\xi-a|<3\sigma\}=0.9974.$$

我们指出，虽然 $N(a,\sigma^2)$ 的密度 $p(x)$ 恒为正，即 ξ 取值于 $\mathbb{R}$ 上任何小区间总有正的概率. 但根据上述计算结果，应用中通常认为 ξ 只取 $(a-3\sigma,a+3\sigma)$ 中的值. 称为 "3σ 原则". 使用这一原则所引起的偏差概率还不到千分之三.

正态分布是最重要的连续型分布. 实践证明，凡是一随机现象是许许多多偶然因素共同作用的总和，各偶然因素所起的作用势均力敌，没有哪个起主导作用. 那么这个随机现象的概率模型就是正态分布. 这一经验主义的看法的确切含义及严格的数学证明称为中心极限定理，我们将在第四章详加介绍.

于是，可以看作 (或者近似地看作) 正态分布的随机变量就广泛存在于客观世界中. 例如，产品的质量指标，诸如尺寸，含量，强度，寿命等等，都可以认为服从正态分布. 这是因为，除了主导因素之外，影响产品质量指标的偶然因素非常多. 例如原材料质地的不均匀，加工时电力的波动，生产者心理上的细微变化，储存方式的差异等等. 这些偶然因素共同作用的结果，使得产品的质量指标呈现为 "两头小中间大" 的正态曲线的形状. 此外，正态分布在误差分析及炮弹弹着点的分布等方面也有相当成功的应用.

2.4.3 伽马分布

设 $\{\xi_t,t\geq 0\}$ 为 2.3.3 节所述普阿松过程. 已经得到 ξ_t 服从参数为 λt 的普阿松分布. 现对自然数 r, 考虑第 r 个质点的到达时间 η_r, 当 $t>0$ 时，它的分布函数为

$$F(t)=P\{\eta_r<t\}=P\{\xi_t\geq r\}=\sum_{k=r}^{\infty}\frac{(\lambda t)^k}{k!}\,e^{-\lambda t}.$$

对 t 求导便得到其密度函数

$$p(t)=F'(t)=\sum_{k=r}^{\infty}\frac{\lambda^k t^{k-1}}{(k-1)!}\,e^{-\lambda t}-\sum_{k=r}^{\infty}\frac{\lambda^{k+1}t^k}{k!}\,e^{-\lambda t}$$

$$= \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t}, \qquad t > 0.$$

回忆伽马函数

$$\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx,$$

它是对一切参数 $r > 0$ 均收敛的广义积分. 特别有

$$\Gamma(n+1) = n\Gamma(n) = \cdots = n!,$$

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} x^{\frac{1}{2}-1} e^{-x} dx = \sqrt{\pi} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \sqrt{\pi}.$$

于是, 普阿松过程中第 r 个质点到达时间 η_r 的密度函数可写为

$$p(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \qquad x > 0. \tag{4.6}$$

将上式中的自然数 r 推广到 r 为正实数情形, 易证 $p(x)$ 满足 (1.20) 式中密度函数的两个条件.

定义 4.3 称以 (4.6) 式中 $p(x)$ 为密度函数的连续型分布为**伽马分布**. 其参数 $\lambda > 0$, $r > 0$. 记作 $\Gamma(\lambda, r)$.

图 4.3 画出了 $\lambda = 1$ 时不同 r 值的伽马分布密度曲线.

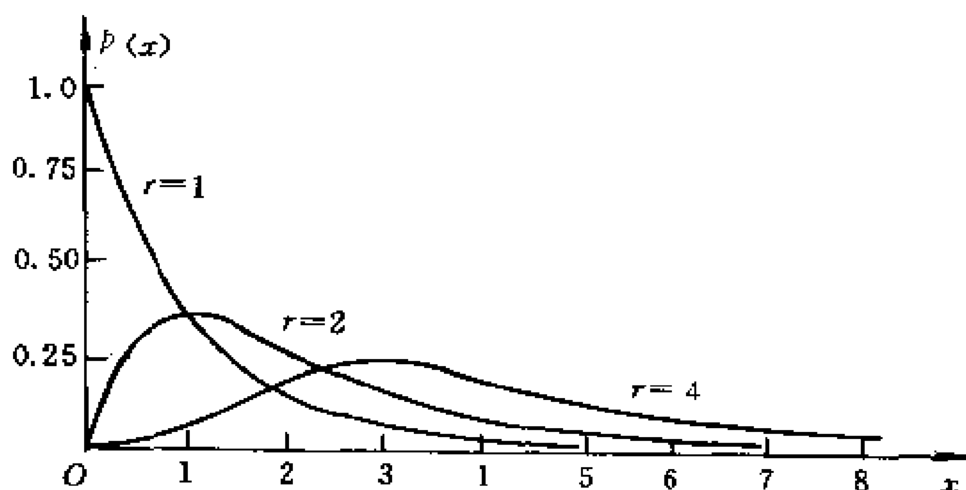


图 4.3 $\Gamma(1, r)$ 密度曲线

与可列重贝努里试验中第 r 次成功等待时间的巴斯卡分布相类似，今后将证明，服从 $\Gamma(\lambda,r)$ 分布的普阿松过程中第 r 个质点到达时间 η_r，可写为质点到达时间间隔的和

$$\eta_r = \tau_1 + \cdots + \tau_r,$$

其中

$$\tau_1 = \eta_1, \quad \tau_k = \eta_k - \eta_{k-1}, \quad 1 < k \leq r.$$

诸 τ_k 有相同的 $\Gamma(\lambda,1)$ 分布，而 $\Gamma(\lambda,1)$ 分布也有相当于贝努里场合中几何分布的相应性质．我们将在下节作进一步研究．

2.4.4 指数分布

定义 4.4 称 $\Gamma(\lambda,1)$ 分布为**指数分布**．其密度函数与分布函数分别为

$$p(x) = \lambda e^{-\lambda x}, \qquad x > 0, \tag{4.7}$$

$$F(x) = 1 - e^{-\lambda x}, \qquad x > 0.$$

由上节分析知，指数分布是普阿松过程中质点到达时间间隔的分布．因此，在应用中常把它作为某些等待时间的概率分布．例如，电子元件的寿命 (即毁坏的等待时间)，排队模型中的服务时间等，都可以用指数分布作为概率模型．

指数分布具有与贝努里场合几何分布相同的无记忆性．可以证明，取非负实值的随机变量 ξ 有指数分布，当且仅当它是无记忆的：

$$P\{\xi > s+t \mid \xi > s\} = P\{\xi > t\}, \qquad 任意\ s,t > 0. \tag{4.8}$$

其证明与 2.2.3 节中的定理 2.1 类似．我们留作习题．这里只说一下 (4.8) 式的含义．假定某种产品的使用寿命 ξ 服从指数分布，那么，(4.8) 式表明，无论它已经被使用了多长一段时间 s，只要还没有损坏，它能再使用一段时间 t 的概率与一件新产品能使用到时间 t 的概率一样．就是说，这种产品将“永远年青”．这正好说明以指数分布作

为寿命分布是有缺陷的. 尽管如此, 在有些场合人们还是愿意采用这种易于计算的分布作为寿命的模型. 当然, 应用中也可以选用正态分布或威布尔分布 (参见附表 I) 作为寿命分布.

例 4.2 假定自动取款机对每位顾客的服务时间 (单位：分钟) 服从 $\lambda = 1/3$ 的指数分布. 如果有一顾客恰好在你前头走到空闲的取款机, 求你 (1) 至少等候 3 分钟； (2) 等候时间在 3 分钟至 6 分钟之间的概率. 如果你到达取款机时, 正有一名顾客使用着收款机, 上述概率又是多少？

解: 以 ξ 表示你前面这位顾客所用服务时间. $F(x)$ 为 ξ 的分布函数. 则所求概率

(1) $P\{\xi > 3\} = 1 - F(3) = e^{-1} = 0.368.$

(2) $P\{3 < \xi < 6\} = F(6) - F(3) = e^{-1} - e^{-2} = 0.233.$

如果你到达时取款机正在为一名顾客服务, 同时没有其他人已在排队等候. 那么由指数分布的无记忆性, 取款机还需花在你前面顾客身上的服务时间, 与他刚到取款机相同, 从而问题的答案不变.

2.4.5 习题

1. 设 ξ 服从区间 $(0,5)$ 上的均匀分布, 求二次方程 $4x^2+4\xi x+\xi+2=0$ 有实根的概率.

2. 假定随机变量 ξ 只取区间 $(0,1)$ 中的值, 且对任何 $0<x<y<1$, ξ 落在子区间 (x,y) 内的概率仅与 $y-x$ 有关. 求证 ξ 服从区间 $(0,1)$ 上的均匀分布.

3. 在 $\triangle ABC$ 内任取一点 M(二维几何概型), 连结 AM 并延长, 与边 BC 相交于 N, 证明点 N 的坐标在线段 BC 构成的区间上均匀分布.

4. 设 ξ 服从 $N(3,4)$ 分布. (1) 求 a 使 $P\{\xi > a\} = 2P\{\xi < a\}$. (2) 求 b 使 $P\{|\xi - 3| < b\} = 0.95$.

5. 在正常的考试中, 学生的成绩应当服从 $N(a,\sigma^2)$ 分布. 如果规定分数在 $a+\sigma$ 以上为“优秀”, a 至 $a+\sigma$ 之间为“良好”, $a-\sigma$

至 a 之间为“一般”, $a-\sigma$ 以下为“较差”. 试求这四个等级的学生各占多大比例.

6. 开汽车从城南到城北火车站. 如果穿行, 则所需时间 (单位: 分钟) 服从 $N(50,100)$ 分布. 若绕行, 则所需时间服从 $N(60,16)$ 分布. 假定现有 (1) 65 分钟; (2) 70 分钟可用, 试计算是穿行还是绕行好些?

7. 已知随机变量 ξ 服从标准正态分布, $\eta=\xi$ 或 $-\xi$, 视 $|\xi|\le 1$ 或 $|\xi|>1$ 而定. 求 η 的分布.

8. 设 ξ 服从参数为 λ 的指数分布, 求 $\eta=[\xi]$ 的分布.

9. 证明非负值随机变量有指数分布的充分必要条件是它有无记忆性 (4.8) 式.

10. 假定一机器的检修时间 (单位: 小时) 是以 $\lambda=1/2$ 为参数的指数分布. 试求 (1) 检修时间超过 2 小时的概率; (2) 若已经修理 4 个小时, 求总共要至少 5 个小时才会修好的概率.

2.5 多维概率分布

2.5.1 随机向量

有些随机试验的结果是多个随机变量. 例如天气预报, 不仅要预报某地某日的最高气温, 还有最低气温, 这就是两个随机变量了. 如果还要预报风力, 湿度, 降水量等, 那么这个随机试验的结果就是多个随机变量. 由于这些随机变量共处于同一随机试验之中, 它们是相互联系的, 因此只进行逐个研究是不够的, 而必须从整体上, 从各分量的相依关系上加以讨论, 这就是随机向量.

定义 5.1 设 $(\Omega,\mathcal{F},P)$ 为一概率空间, 对于 $i=1,\cdots,n$, $\xi_i(\omega)$ 为其上随机变量, 即

$$\{\omega:\xi_i(\omega)\in B_i\}\in\mathcal{F},\qquad 任\quad B_i\in\mathcal{B}, \tag{5.1}$$

则称 $\vec{\xi}(\omega)=(\xi_1(\omega),\cdots,\xi_n(\omega))$ 为这个概率空间上的**随机向量**, 又称作 n 维随机变量.

n 维随机向量取值于 $\Re^n$, 它是一维直线 $\Re = (-\infty, +\infty)$ 的 n 维笛卡儿乘积. 其上波雷尔集类 $\mathcal{B}^n$ 也是一维波雷尔集类 $\mathcal{B}$ 的笛卡儿乘积 (参见 1.6.4 节), 即 $\mathcal{B}^n = \sigma(\mathcal{P})$, 其中 $\mathcal{P}$ 为形如 $B_1 \times \cdots \times B_n$ 的可测矩形全体构成的类. 下面给出随机向量定义 (5.1) 式的两种等价形式.

定理 5.1 $(\Omega, \mathcal{F}, P)$ 上 n 维向量值函数 $\vec{\xi}(\omega) = (\xi_1(\omega), \cdots, \xi_n(\omega))$ 为随机向量, 当且仅当如下条件之一成立:

(1) $\{\omega : \vec{\xi}(\omega) \in B\} \in \mathcal{F}$, 任 $B \in \mathcal{B}^n$; hfill (5.2)

(2) $\{\omega : \xi_1(\omega) < x_1, \cdots, \xi_n(\omega) < x_n\} \in \mathcal{F}$. 任 $x_i \in \Re$. (5.3)

证明: (5.1)$\Longrightarrow$(5.2)

记 $\mathcal{E} = \{B \subset \Re^n : \xi^{-1}(B) \in \mathcal{F}\}$, 易证 $\mathcal{E}$ 为 σ 代数. (5.1) 式表明 $\mathcal{E} \supset \mathcal{P} = \{B_1 \times \cdots \times B_n : B_i \in \mathcal{B}, i = 1, \cdots, n\}$, 故 $\mathcal{E} \supset \sigma(\mathcal{P}) = \mathcal{B}^n$. 即有 (5.2) 式成立.

(5.2)$\Longrightarrow$(5.3)

取 (5.2) 中 $B = (-\infty, x_1) \times \cdots \times (-\infty, x_n) \in \mathcal{B}^n$ 即得.

(5.3)$\Longrightarrow$(5.1)

对每个固定的 i, 当 $j \neq i$ 时取 $x_j = m$, 则由 (5.3) 知

$$E_m = \{\omega : \xi_1 < m, \cdots, \xi_{i-1} < m, \xi_i < x_i, \xi_{i+1} < m, \cdots, \xi_n < m\} \in \mathcal{F},$$

从而对于 $i = 1, \cdots, n$ 有

$$\{\omega : \xi_i(\omega) < x_i\} = \bigcup_{m=1}^{\infty} E_m \in \mathcal{F}, \qquad \text{任} \quad x_i \in \Re.$$

再由 2.1.1 节的定理 1.2 知对一切 i (5.1) 式成立. 定理证毕.

有关 n 维随机向量的讨论, 许多内容都与 1 维情形相类似, 我们只作简略叙述, 重点放在多维情形的特异之处. 另一方面, 为叙述简便, 下面仅以 $n = 2$ 维为例, 所得结果适用于高维随机变量. 如有例外将另加说明.

2.5.2 联合分布

定义 5.2 设 (ξ,η) 为 $(\Omega,\mathcal{F},P)$ 上随机向量，称

$$F(x,y)=P\{\xi<x,\eta<y\},\qquad x,y\in\Re \tag{5.4}$$

为 (ξ,η) 的**联合分布函数**.

如下定理给出联合分布函数的本征性质.

定理 5.2 $F(x,y)$ 有性质

1° $F(x,y)$ 对每个自变量都是单调非降的;

2° $F(x,y)$ 对每个自变量都是左连续的;

3° $\lim\limits_{x\to-\infty}F(x,y)=\lim\limits_{y\to-\infty}F(x,y)=1-\lim\limits_{\substack{x\to+\infty\\ y\to-\infty}}F(x,y)=0$;

4° $F(x,y)$ 在任一矩形 $[a_1,b_1)\times[a_2,b_2)$ 上增量 ΔF 非负. 其中

$$\Delta F=F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2).$$

证明: 只需证 4°. 由 (5.4) 式可知, ΔF 恰为 (ξ,η) 落入所述矩形中的概率 $P\{a_1\le\xi<b_1,a_2\le\eta<b_2\}$, 自然是非负的. 证毕.

在 $n=1$ 维情形. 分布函数在区间上增量非负是性质 1° 的推论. 但对 $n\ge2$ 维情形, 性质 4° 不能由 $1^\circ,2^\circ,3^\circ$ 推出. 因而 4° 是多维分布函数特有的性质. 且看如下的例.

例 5.1 设 $F(x,y)=1$, 当 $x+y>0$; $=0$, 否则. 不难验证此 $F(x,y)$ 有性质 $1^\circ,2^\circ,3^\circ$, 但在 4° 中取 $a_1=a_2=0$, $b_1=b_2=1$, 则有

$$\Delta F=F(1,1)-F(0,1)-F(1,0)+F(0,0)=-1<0.$$

也可从另一角度引进联合分布函数 $F(x,y)$, 由 (5.2) 知随机向量 (ξ,η) 是 $(\Omega,\mathcal{F},P)$ 到相空间 $(\Re^2,\mathcal{B}^2)$ 中的可测变换, 从而可在 $\mathcal{B}^2$ 上引出一概率测度

$$\mathbf{F}(B)=P\{\omega:(\xi(\omega),\eta(\omega))\in B\},\qquad 任\quad B\in\mathcal{B}^2. \tag{5.5}$$

称此 $\mathbf{F}$ 为 (ξ,η) 的**联合分布**. 通过 (5.5) 式可用联合分布 $\mathbf{F}$ 回答有关 (ξ,η) 的任何概率问题. 特别, 联合分布函数 $F(x,y)$ 就是 $\mathbf{F}$ 在 $B=(-\infty,x)\times(-\infty,y)$ 上的值.

与 1 维情形类似, (5.5) 式的左方就是 B 上关于 $F(x,y)$ 的勒贝格 - 斯蒂尔吉斯积分, 即有

$$P\{(\xi,\eta)\in B\}=\int\int_B dF(x,y),\qquad 任\quad B\in\mathcal{B}^2. \tag{5.6}$$

不熟悉这种积分的读者, 只要掌握上式在离散型与连续型情形的具体形式, 即下面的 (5.8) 式与 (5.10) 式即可.

定义 5.3 若随机向量 (ξ,η) 有至多可列对可能值 (x_i,y_j), $i,j=1,2,\cdots$, 则称此随机向量及其联合分布为**离散型**的. 而

$$p_{ij}=P\{\xi=x_i,\eta=y_j\},\qquad i,j=1,2,\cdots \tag{5.7}$$

称为 (ξ,η) 的**联合分布**.

易见, 离散型联合分布 $\{p_{ij}\}$ 满足 $p_{ij}\geq 0$ 及 $\sum\limits_{i,j}p_{ij}=1$. 并且, 此时 (5.6) 式化为

$$P\{(\xi,\eta)\in B\}=\sum_{(x_i,y_j)\in B}p_{ij},\qquad 任\quad B\in\mathcal{B}^2. \tag{5.8}$$

例 5.2 将两个不同的球任意放入编号为 1,2,3 的三个盒, 每球入各盒均等可能. 以 ξ 表示空盒个数, η 表示有球盒的最小编号. 试求 (ξ,η) 的联合分布.

解: ξ 的可能值有 1,2, 而 η 的可能值为 1, 2, 3. 由于每个球有 3 种放法, 故样本点总数为 9. 于是对一切 j 有 $p_{2j}=P\{$ 两球均在 j 号盒中 $\}=1/9$. 其次, $p_{12}=P\{1$ 号盒空, 2,3 号盒各一球 $\}=2/9$. 再者, $p_{11}=P\{1,2$ 号盒各一球 $\}+P\{1,3$ 号盒各一球 $\}=4/9$. 最后 $p_{23}=P\{\emptyset\}=0$. 用 (5.8) 式容易求得有关 (ξ,η) 的概率, 例如我们有

$$P\{\xi=\eta\}=\sum_{x_i=y_j}p_{ij}=p_{11}+p_{22}=\frac{5}{9}.$$

可以象一维离散型那样，将所得到 (ξ,η) 的联合分布写作矩阵

$$\begin{pmatrix} (1,1) & (1,2) & (2,1) & (2,2) & (2,3) \\ 4/9 & 2/9 & 1/9 & 1/9 & 1/9 \end{pmatrix}.$$

应用中也时常把二维离散型分布列成如下形式的表：

p_{ij}		y_j: 1	2	3	$p_{i\cdot}$
x_i	1	4/9	2/9	0	6/9
	2	1/9	1/9	1/9	3/9
$p_{\cdot j}$		5/9	3/9	1/9	1

(5.8) 式的一个特例是，联合分布函数 $F(x,y)$ 也可通过 $\{p_{ij}\}$ 表出，即有

$$F(x,y) = \sum_{x_i<x \text{且} y_j<y} p_{ij}, \qquad (x,y) \in \Re^2.$$

将此式中的求和换为积分，可类比地得到连续型的定义.

定义 5.4 设随机向量 (ξ,η) 的联合分布函数

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(u,v)\, dudv, \qquad (x,y) \in \Re^2, \tag{5.9}$$

其中被积函数满足

$$p(x,y) \geq 0 \quad \text{且} \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x,y)\, dxdy = 1,$$

则称 (ξ,η) 及其概率分布为**连续型**的，称 $p(x,y)$ 为其**联合密度函数**.

对于连续型随机向量，(5.6) 式化为

$$P\{(\xi,\eta) \in B\} = \int\int_B p(x,y)\, dxdy, \qquad \text{任} \quad B \in \mathcal{B}^2. \tag{5.10}$$

例 5.3 考虑

$$p(x,y) = x+y, \qquad 0 < x,y < 1, \tag{5.11}$$

易见 $p(x,y) \geq 0$ 且

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\,dxdy = \int_0^1\int_0^1 (x+y)\,dxdy = \frac{1}{2}+\frac{1}{2} = 1.$$

即它可作为连续型的 2 维联合密度函数. 设随机向量 (ξ,η) 以此 $p(x,y)$ 为联合密度函数, 用 (5.10) 式可得

$$P\{\xi+\eta<1\} = \int\int_{x+y<1} p(x,y)dxdy$$

$$= \int\int_{x,y>0\text{且}x+y<1} (x+y)\,dxdy = \int_0^1 dx\int_0^{1-x}(x+y)\,dy = \frac{1}{3}.$$

2.5.3 边缘分布

本节考虑随机向量的分量各自的概率分布. 在 (ξ,η) 的联合分布函数 (5.4) 式中, 令 $y\uparrow+\infty$, 则事件 $\{\eta<y\}\uparrow\Omega$. 利用概率的下连续性便得 ξ 的分布函数

$$F_1(x) = P\{\xi<x\} = \lim_{y\to+\infty} F(x,y) = F(x,+\infty), \qquad x\in\Re. \quad (5.12)$$

称它为 ξ 的**边缘分布函数**. 同理 η 的边缘分布函数为

$$F_2(y) = P\{\eta<y\} = \lim_{x\to+\infty} F(x,y) = F(+\infty,y), \qquad y\in\Re. \quad (5.13)$$

"边缘" 一词来源于离散型情形. 在 2 维离散型概率分布 $\{p_{ij}\}$ 的列表表示中 (参见例 5.2 的表), 将各行求和写在表的最右一列, 再将各列求和写在表的最下一行. 由于

$$p_{i\cdot} = \sum_j p_{ij} = \sum_j P\{\xi=x_i,\eta=y_j\} = P\{\xi=x_i\}, \qquad (5.14)$$

$$p_{\cdot j} = \sum_i p_{ij} = \sum_i P\{\xi=x_i,\eta=y_j\} = P\{\eta=y_j\}, \qquad (5.15)$$

因此，位于表中右边与下边的数列分别是 ξ 与 η 各自的分布. 故称为边缘分布. 例 5.2 的表中已算出 ξ 与 η 的边缘分布分别是

$$\begin{pmatrix} 1 & 2 \\ 6/9 & 3/9 \end{pmatrix} \quad 与 \quad \begin{pmatrix} 1 & 2 & 3 \\ 5/9 & 3/9 & 1/9 \end{pmatrix}.$$

对连续型随机向量 (ξ,η), 在 (5.9) 式中令 $y \to +\infty$ 得 ξ 的边缘分布函数

$$F_1(x) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} p(u,v)\, dudv = \int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} p(u,v)dv\right] du,$$

这里已将 $F_1(x)$ 写成变上限积分形式. 故分量 ξ 仍为连续型. 有边缘密度函数

$$p_1(x) = \int_{-\infty}^{+\infty} p(x,y)\, dy, \qquad x \in \Re. \tag{5.16}$$

同理，分量 η 也是连续型的，其边缘密度函数为

$$p_2(y) = \int_{-\infty}^{+\infty} p(x,y)\, dx, \qquad y \in \Re. \tag{5.17}$$

下面计算例 5.3 中 ξ,η 的边缘密度. 当 $x \in (0,1)$ 时

$$p_1(x) = \int_0^1 (x+y)\, dy = x + \frac{1}{2}. \tag{5.18}$$

而对于其它的 x, (5.13) 式中被积函数恒为 0, 故 $p_1(x) = 0$. 对称可得 η 的边缘密度

$$p_2(y) = y + \frac{1}{2}, \qquad 0 < y < 1. \tag{5.19}$$

由 (5.12) 至 (5.17) 式可见，分量的边缘分布由联合分布完全确定. 但是逆命题不真，即边缘分布不能确定联合分布.

例 5.4 设

$$q(x,y) = (x+\frac{1}{2})(y+\frac{1}{2}), \qquad 0 < x,y < 1.$$

此 $q(x,y)$ 恰为由 (5.18) 与 (5.19) 给出的 $p_1(x)$ 与 $p_2(y)$ 之乘积. 故易知 $q(x,y)$ 是二元密度函数, 且以 $p_1(x)$ 与 $p_2(y)$ 为其边缘密度函数. 但是 $q(x,y)$ 与 (5.11) 式给出的 $p(x,y)$ 显然不同, 即联合分布不唯一.

为强调这一点, 我们再看一个离散型的例子.

例 5.5 袋中有 5 张外形相同的卡片, 其中 3 张写上数字 0, 另 2 张写着数字 1. 现从袋中任取出两张卡片, 分别以 ξ,η 表第 1 张与第 2 张卡片上的数字. 那么对有放回与不放回两种方式, 可以分别得到 (ξ,η) 的联合分布如下:

有放回方式

p_{ij}	0	1	$p_{i\cdot}$
0	9/25	6/25	3/5
1	6/25	4/25	2/5
$p_{\cdot j}$	3/5	2/5	1

不放回方式

p_{ij}	0	1	$p_{i\cdot}$
0	6/20	6/20	3/5
1	6/20	2/20	2/5
$p_{\cdot j}$	3/5	2/5	1

两种方式中 ξ 的边缘分布相同, η 的边缘分布也相同, 但是 (ξ,η) 的联合分布却不同. 即边缘分布不能确定联合分布.

相同的边缘分布可构成不同的联合分布, 这反映出两个分量的结合方式不同, 相依程度不同. 在上述例 5.5 中, 有放回方式 ξ 与 η 关系较松散, 而不放回方式 ξ 与 η 关系较密切. 相依程度这种差异在各自的边缘分布中没有表现, 因此, 单纯地逐个研究各分量的分布是不够的, 而必须把它们放在一起考察其联合分布.

对于 $n>2$ 的高维情形, $(\xi_1,\cdots,\xi_n)$ 的任何子向量 $(\xi_{i_1},\cdots,\xi_{i_k})$ $(1\le k\le n-1)$ 的分布称作 k 维边缘分布. 假定 $(\xi_1,\cdots,\xi_n)$ 有联合密度函数 $p(x_1,\cdots,x_n)$, 我们只需将它对与指标 $i_1,\cdots,i_k$ 无干的 x_i 的一切可能值取积分, 便可得到相应的 k 维边缘密度函数. 离散型情形只需把积分换为求和.

2.5.4 均匀分布与正态分布

本节介绍两种最常见的 2 维连续型分布．描述 2 维几何概型的随机向量就服从 2 维均匀分布．

定义 5.5 设 $D \in \mathcal{B}^2$ 满足 $0 < m(D) < +\infty$, 称以

$$p(x, y) = \frac{1}{m(D)}, \qquad (x, y) \in D \tag{5.20}$$

为密度函数数的连续型分布为 D 上的**均匀分布**.

若 (ξ, η) 服从 D 上的均匀分布，则运用 (5.10) 式可得，对任何 $B \subset D$ 有

$$P\{(\xi, \eta) \in B\} = \int\int_B p(x, y) dx dy = \frac{m(B)}{m(D)},$$

即 (ξ, η) 落入 D 之任何波雷尔子集 B 内的概率与 B 的勒贝格测度 $m(B)$ 成正比，而与 B 的形状及位置等无关.

例 5.6 设 (ξ, η) 服从单位圆 $D = \{(x, y) : x^2 + y^2 < 1\}$ 上的均匀分布，试求它的边缘密度函数.

解: (ξ, η) 的联合密度函数为

$$p(x, y) = \frac{1}{\pi}, \qquad x^2 + y^2 < 1.$$

我们用 (5.16) 式求 ξ 的边缘密度函数．当 $|x| \geq 1$ 时，$p(x, y) = 0$ 故 $p_1(x) = 0$; 而当 $|x| < 1$ 时

$$p_1(x) = \int_{-\infty}^{+\infty} p(x, y)\, dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy = \frac{2}{\pi}\sqrt{1 - x^2}.$$

对称可得 η 的边缘密度函数为

$$p_2(y) = \frac{2}{\pi}\sqrt{1 - y^2}, \qquad -1 < y < 1.$$

注意单位圆上均匀分布的边缘分布不是 1 维均匀分布.

定义 5.6 设参数 $a_1, a_2, \sigma_1, \sigma_2, r$ 满足 $\sigma_1 > 0, \sigma_2 > 0$ 及 $-1 < r < 1$. 称以

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right]\right\} \tag{5.21}$$

为密度函数的连续型分布为**2 维正态分布**.

注意 $p(x,y) > 0$ 显然, 而作为 2 维密度函数的另一条件

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)dxdy = 1$$

则留待后面验证. 2 维正态分布的密度 $p(x,y)$ 的图形是一个钟形曲面. 图 5.1 画出了它的一部分的示意图. 为计算 2 维正态的边缘密度函数, 我们指出, 有关正态分布的积分运算往往要采用配方, 换元等手法. 于是据 (5.16) 式可算出 2 维正态分布的边缘密度函数

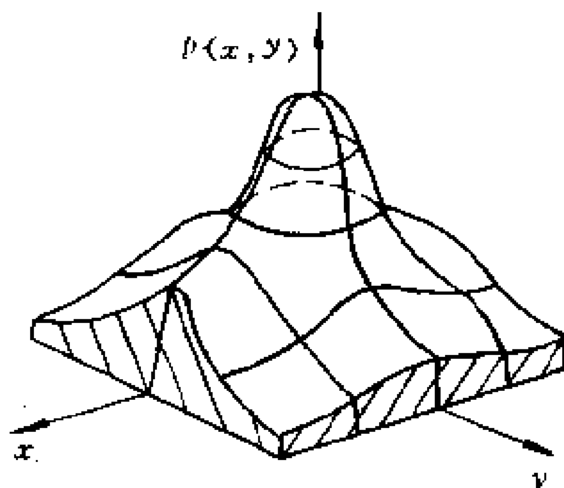


图 5.1 二维正态密度曲面

$$p_1(x) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2r(x-a_1)(b-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right]\right\} dy$$

$$= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sqrt{1-r^2}} \exp\left\{-\frac{u^2-2ruv+v^2}{2(1-r^2)}\right\} dv$$

$$= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2}\left[(\frac{v-ru}{\sqrt{1-r^2}})^2 + u^2\right]\right\} dv$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma_1^2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma_1^2}}$$

即 ξ 服从 $N(a_1, \sigma_1^2)$ 分布. 同理可得 η 服从 $N(a_2, \sigma_2^2)$ 分布. 至此知 2 维正态分布的边缘分布是 1 维正态分布, 这一点与圆上均匀分布情形不同. 注意到 2 维正态分布的两个边缘密度 $p_1(x)$ 与 $p_2(x)$ 中都不含参数 r, 于是我们再次得到边缘分布不能确定联合分布的例子. 同是以 $N(a_1, \sigma_1^2)$ 与 $N(a_2, \sigma_2^2)$ 作为边缘分布, 参数 r 在区间 $(-1,1)$ 内任取, 可得到不同的联合正态分布 (5.21). 今后我们会看到, 参数 r 确实反映了两个分量 ξ 与 η 的相依程度.

回忆在 1 维正态分布中已经证得 $I = \int_{-\infty}^{+\infty} p_1(x)dx = 1$. 化重积分为累次积分便补证了

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)\,dxdy = \int_{-\infty}^{+\infty} p_1(x)\,dx = 1.$$

2.5.5 习题

1. 甲从 1,2,3,4 中任取一数 ξ, 乙再从 $1,\cdots,\xi$ 中任取一数 η. 试求 (ξ,η) 的联合分布与边缘分布.

2. 甲乙二人轮流投篮. 假定每次甲的命中率为 0.4, 乙的命中率为 0.6, 且各次投篮相互独立. 现甲先投, 乙再投, 直至有人命中为止. 试求甲与乙投篮次数 ξ 与 η 的联合分布与边缘分布.

3. 设 (ξ,η) 的联合密度函数为

$$p(x,y) = \frac{1}{2}\sin(x+y), \qquad 0 < x, y < \frac{\pi}{2},$$

试求 (1) (ξ,η) 的联合分布函数; (2) η 的边缘密度函数.

4. 设 $f(x)$ 是某非负值随机变量的密度函数, 试证

$$p(x,y) = \frac{f(x+y)}{x+y}, \qquad x, y > 0$$

是 2 维密度函数.

5. 设分布函数 $F_1(x)$ 与 $F_2(x)$ 对应的密度函数为 $p_1(x)$ 与 $p_2(x)$. 证明对任何 $\alpha \in (-1,1)$,

$$p_\alpha(x,y) = p_1(x)p_2(y)\{1 + \alpha[2F_1(x)-1][2F_2(y)-1]\}$$

是 2 维密度函数，且以 $p_1(x)$ 与 $p_2(y)$ 为其边缘密度函数.

6. 设 (ξ,η) 的联合密度函数

$$p(x,y)=c\,xy^2,\qquad 0<x<2,\ 0<y<1.$$

试求 (1) 常数 c; (2) ξ,η 至少有一个小于 1/2 的概率.

7. 在可列重贝努里试验中，以 ξ_i 表第 i 次成功的等待时间，试求 (ξ_1,ξ_2) 的 (1) 联合分布； (2) 边缘分布.

8. 设 (ξ,η) 服从区域 $D=\{(x,y):0<x^2<y<x<1\}$ 上的均匀分布. 试求 (1)ξ 的边缘密度函数; (2)$P\{\eta>1/2\}$.

9. 设 (ξ,η) 为 2 维正态随机向量，求 (ξ,η) 落入区域

$$D=\left\{(x,y):\frac{(x-a_1)^2}{\sigma_1^2}-\frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}\le\lambda^2\right\}$$

内的概率.

10. 用概率思想证明，对任意 $a>0$ 有

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{-\frac{x^2}{2}}dx\le\sqrt{1-e^{-a^2}}.$$

11. 设随机向量 (ξ,η,ζ) 有联合密度函数

$$p(x,y)=xze^{-(x+xy+z)},\qquad x,y,z>0,$$

试求 (1) ξ,η 的 1 维边缘密度; (2) (ξ,ζ) 的 2 维边缘密度.

2.6 随机变量的独立性

2.6.1 条件分布

设 $(\Omega,\mathcal{F},P)$ 为一概率空间. 任固定 $B\in\mathcal{F},P(B)>0$. 则 $\mathcal{F}$ 中的每一事件 A 有关于事件 B 的条件概率 $P(A|B)$. 特别，此概率空间中的随机变量 ξ 关于这个条件概率的分布就是已知事件 B 已发生后 ξ 的条件分布. 比如，ξ 关于 B 的条件分布函数

$$F(x|B)=P\{\xi<x|B\}=\frac{P\{\xi<x,B\}}{P(B)},\qquad x\in\Re.\tag{6.1}$$

现假定 $(\Omega, \mathcal{F}, P)$ 上还有另一随机变量 η, 那么当 $B=\{\eta=y\}$ 有正概率时，由 (6.1) 式定义的 $F(x|\eta=y)$ 就是已知 $\eta=y$ 时 ξ 的条件分布函数.

对于 (ξ,η) 为离散型随机向量的场合，假定 $\{p_{ij}\}$ 为其联合分布，那么对于使 η 的边缘分布 $p_{\cdot j}=P\{\eta=y_j\}>0$ 的 j, 我们有

$$P\{\xi=x_i|\eta=y_j\}=\frac{p_{ij}}{p_{\cdot j}}, \qquad i=1,2,\cdots. \tag{6.2}$$

定义 6.1 对以 $\{p_{ij}\}$ 为联合分布的离散型随机向量 (ξ,η), 当 $p_{\cdot j}>0$ 时，称 (6.2) 式定义的 $P\{\xi=x_i|\eta=y_j\}, i=1,2,\cdots$ 为已知 $\eta=y_j$ 时 ξ 的**条件分布**.

对称可定义已知 $\xi=x_i$ 时 η 的条件分布 $P\{\eta=y_j|\xi=x_i\}, j=1,2,\cdots$.

容易验证，(6.2) 式定义的条件分布满足离散型概率分布的两个条件 $P\{\xi=x_i|\eta=y_j\}\geq 0$ 与 $\sum\limits_i P\{\xi=x_i|\eta=y_j\}=1$.

由定义立得，在上节例 5.2 中，已知 $\xi=1$ 时 η 的条件分布为

$$P\{\eta=1|\xi=1\}=\frac{4/9}{6/9}=\frac{2}{3}, \qquad P\{\eta=2|\xi=1\}=\frac{2/9}{6/9}=\frac{1}{3}.$$

而已知 $\xi=2$ 时，η 的条件分布是 1,2,3 这三点上的等可能分布.

不难验证，在例 5.5 中的有放回情形，条件分布与无条件 (边缘) 分布相同. 就是说，第 1 次取数的结果对第 2 次取出数的分布没有影响. 但是对于不放回情形，则条件分布不同于无条件分布. 这与我们的直观看法是一致的.

下面考察连续型随机向量 (ξ,η). 由于 η 取任何指定值 y 的概率为 0, 故不能直接用 (6.2) 式作为条件分布的定义. 但如下的讨论表明，与 (6.2) 式相当的密度函数形式仍可以作为条件分布的定义.

假设 (ξ,η) 有联合密度函数 $p(x,y)$, 而 η 的边缘密度函数为 $p_2(y)$. 对使事件 $B=\{\eta\in[y,y+\Delta y)\}$ 有正概率的 y 与 $\Delta y>0$, ξ 关于 B 的条件分布函数为

$$F(x|B)=P\{\xi<x|\eta\in[y,y+\Delta)\}$$

$$= \frac{P\{\xi < x, \eta \in [y, y+\Delta y)\}}{P\{\eta \in [y, y+\Delta y)\}} = \frac{\int_{-\infty}^{x}\int_{y}^{y-\Delta y} p(u,v)\, dudv}{\int_{y}^{y+\Delta y} p_2(v)\, dv}.$$

利用积分中值定理，存在 $y', y'' \in [y, y+\Delta y)$ 使

$$F(x|\eta \in [y, y+\Delta y)) = \int_{-\infty}^{x} \frac{p(u,y')}{p_2(y'')}\, du.$$

令 $\Delta y \to 0$, 如果上式极限存在，则应有

$$F(x|\eta = y) = \int_{-\infty}^{x} \frac{p(u,y)}{p_2(y)}\, du.$$

这表明已知 $\eta = y$ 时 ξ 的条件分布仍为连续型，其密度函数为

$$p(x|\eta = y) = \frac{p(x,y)}{p_2(y)}, \qquad x \in \Re. \tag{6.3}$$

定义 6.2 假设随机向量 (ξ, η) 有联合密度 $p(x,y)$, 如果 y 使 $p_2(y) > 0$, 则称由 (6.3) 式定义的 $p(x|\eta = y)$ 为已知 $\eta = y$ 时 ξ 的**条件密度函数**.

对称可定义已知 $\xi = x$ 时 η 的条件密度函数 $p(y|\xi = x)$.

不难验证，条件密度满足密度函数的两个条件.

回到上节例 5.6, (ξ, η) 服从单位圆上均匀分布，当 $|y| < 1$ 时 $p_2(y) > 0$, 故已知 $\eta = y \in (-1, 1)$ 时 ξ 的条件密度函数为

$$p(x|\eta = y) = \frac{1}{2\sqrt{1-y^2}}, \qquad 当\ -\sqrt{1-y^2} < x < \sqrt{1-y^2},$$

即 ξ 的条件分布是区间 $(-\sqrt{1-y^2}, \sqrt{1-y^2})$ 上的均匀分布.

最后我们指出，2 维正态随机向量的条件分布是 1 维正态分布. 据 (6.3) 式直接计算可得，如果 (ξ, η) 的联合密度函数由上节 (5.21) 式给出，那么已知 $\eta = y$ 时 ξ 的条件分布为

$$N\left(a_1 + r\frac{\sigma_1}{\sigma_2}(y - a_2),\ \sigma_1^2(1-r^2)\right) \tag{6.4}$$

分布. 如果参数 $r=0$, 则 (6.4) 式给出的条件分布化为 $N(a_1,\sigma_1^2)$ 分布. 即 ξ 的条件分布与无条件 (边缘) 分布相同. 此时, 2 维正态的联合密度函数恰为两个边缘密度之积. 在分析离散型的例 5.5 时已说过, 条件分布与无条件分布相同, 表明随机向量的两个分量取值互不影响, 这正是随机变量的独立性.

2.6.2 相互独立的随机变量

回忆随机事件相互独立是指其中一事件发生与否对其它事件发生的概率无影响, 或者说, 条件概率等于无条件概率. 随机变量的独立性是类似的. 上节讨论过的有放回情形两次取数的结果以及 $r=0$ 时 2 维正态随机变量, 都具有条件分布与无条件分布相同这一特点. 这就是独立随机变量的例子.

定义 6.3 设 $\xi_1,\cdots,\xi_n$ 是概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 个随机变量, 如果它们的联合分布函数等于各自边缘分布函数之积, 即

$$F(x_1,\cdots,x_n)=F_1(x_1)\cdots F_n(x_n),\qquad x_i\in\Re,\ i=1,\cdots,n. \tag{6.5}$$

则称 $\xi_1,\cdots,\xi_n$**相互独立**. 一族无限多个随机变量称作相互独立的, 如果其中任意有限个相互独立.

若 $\xi_1,\cdots,\xi_n$ 独立, 考虑其任意部分随机变量组成的子向量. 在 (6.5) 式中令与子向量无干的所有 $x_i\to+\infty$, 则左方化为子向量的边缘分布函数, 右方相应地化为子向量各分量的边缘分布函数之积. 于是我们得到

定理 6.1 如果随机变量 $\xi_1,\cdots,\xi_n$ 相互独立, 则其中任何一部分随机变量仍独立.

此定理的逆不真. 反例见习题.

以下仍以 $n=2$ 为例讨论独立性定义 (6.5) 式的种种等价形式, 所得结果适用于 $n>2$ 的情形. 读者可根据需要选用适当的等价形式.

定理 6.2 随机变量 (ξ,η) 相互独立，当且仅当

$$P\{\xi\in B_1,\eta\in B_2\}=P\{\xi\in B_1\}P\{\eta\in B_2\},\qquad 任\quad B_1,B_2\in\mathcal{B}. \tag{6.6}$$

证明: (6.5) 式仅仅是 (6.6) 式中 $B_1=(-\infty,x_1),B_2=(-\infty,\ x_2)$ 的特殊情形. 故 (6.6) 的充分性不待证. 下证 (6.6) 式是必要的. 先固定 x_2, 记

$$\mathcal{E}=\{B_1:P(\xi\in B_1,\eta<x_2)=P(\xi\in B_1)P(\eta<x_2)\}.$$

则由 (6.5) 式知 $\mathcal{E}\supset\mathcal{P}=\{(-\infty,x_1):x_1\in\Re\}$. 不难验证 $\mathcal{E}$ 为 σ 代数, 故 $\mathcal{E}\supset\sigma(\mathcal{P})=\mathcal{B}$. 这就证明了

$$P\{\xi\in B_1,\eta<x_2\}=P\{\xi\in B_1\}P\{\eta<x_2\},\quad 任\quad B_1\in\mathcal{B},x_2\in\Re. \tag{6.7}$$

再固定 $B_1\in\mathcal{B}$, 记

$$\mathcal{G}=\{B_2:\ \ (6.6)\ 式成立\}.$$

同样有 $\mathcal{G}\supset\mathcal{P}$ 且 $\mathcal{G}$ 为 σ 代数, 故 $\mathcal{G}\supset\sigma(\mathcal{P})=\mathcal{B}$. 至此 (6.6) 式的必要性得证.

定理 6.3 若 (ξ,η) 为离散型随机向量, 则 ξ 与 η 独立的充分必要条件是它的联合分布等于边缘分布之积, 即

$$p_{ij}=p_{i\cdot}p_{\cdot j},\qquad 任\quad i,j=1,2,\cdots. \tag{6.8}$$

证明: (6.6) 式中取 $B_1=\{x_i\},B_2=\{y_j\}$, 便得 (6.8) 式成立. 反过来说, 如果有 (6.8) 式成立, 则对任何 $B_1,B_2\in\mathcal{B}$ 有

$$\begin{aligned}P\{\xi\in B_1,\eta\in B_2\}&=\sum_{x_i\in B_1}\sum_{y_j\in B_2}p_{ij}=\sum_{x_i\in B_1}\sum_{y_j\in B_2}p_{i\cdot}p_{\cdot j}\\&=\sum_{x_i\in B_1}p_{i\cdot}\sum_{y_j\in B_2}p_{\cdot j}=P\{\xi\in B_1\}P\{\eta\in B_2\}.\end{aligned}$$

即 (6.6) 式为真. 定理证完.

定理 6.4 若 (ξ,η) 为连续型随机变量，则 ξ 与 η 相互独立的充分必要条件是联合密度函数等于边缘密度函数之积. 即

$$p(x,y)=p_1(x)p_2(y),\qquad \text{对几乎一切 } x,y\in\Re. \tag{6.9}$$

这里，"几乎一切"意即允许有勒贝格零测度集的例外. 如前所述，密度函数在零测集上的差异不影响概率分布.

证明: 如果 (6.9) 式成立，两边取变上限积分立得 (6.5) 式，反过来说，由 (6.5) 式可得

$$\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)\,dudv=\int_{-\infty}^{x}\int_{-\infty}^{y}p_1(u)p_2(v)\,dudv,$$

两边对 x,y 各取一次导数 (绝对连续的分布函数几乎处处可导) 便得到 (6.9) 式. 证毕.

利用定理 6.3 可知，上节例 5.5 中有放回情形的 ξ 与 η 是相互独立的，不放回情形则不独立. 再利用定理 6.4 知，以上节例 5.4 中 $q(x,y)$ 为联合密度的随机变量相互独立，而以例 5.3 中 $p(x,y)$ 为联合密度的随机变量则不独立. 此外，2 维正态随机变量相互独立的充分必要条件是参数 $r=0$.

最后我们来证明，独立随机变量的函数仍独立.

定理 6.5 设随机变量 ξ_1 与 ξ_2 相互独立，则对任意的波雷尔函数 f_1 与 f_2,$\eta_1=f_1(\xi_1)$ 与 $\eta_2=f_2(\xi_2)$ 仍独立.

证明: 由 2.1.1 节的定理 1.3 知 η_1 与 η_2 均为随机变量. 对 $i=1,2$, 记 $A_i=f_i^{-1}(B_i)=\{x:f_i(x)\in B_i\}$, 则因 f_i 为波雷尔函数，故当 $B_i\in\mathcal{B}$ 时 $A_i\in\mathcal{B}$. 于是有

$$P\{\eta_1\in B_1,\eta_2\in B_2\}=P\{f_1(\xi_1)\in B_1,f_2(\xi_2)\in B_2\}$$

$$=P\{\xi_1\in A_1,\xi_2\in A_2\}=P\{\xi_1\in A_1\}P\{\xi_2\in A_2\}$$

$$=P\{f_1(\xi_1)\in B_1\}P\{f_2(\xi_2)\in B_2\}=P\{\eta_1\in B_1\}P\{\eta_2\in B_2\}.$$

由 (6.6) 式知 η_1 与 η_2 相互独立.

例 6.1 假定在一段确定时间内，放射性物质发射出的 α 粒子数服从参数为 λ 的普阿松分布. 如果每个发射出的 α 粒子被记录的概率均为 p, 且各粒子能否被记录相互独立，求证在这段时间内被记录下的 α 粒子数 ξ 与未被记录的 α 粒子数 η 相互独立.

解: 由于普阿松分布在随机选择下不变 (参见例 3.2), 故 ξ, η 均服从普阿松分布，参数分别为 λp 与 λq $(q = 1 - p)$. 于是，对任何非负整数 m 与 n 有

$$
\begin{aligned}
&P\{\xi=m, \eta=n\}\\
&= P\{\xi+\eta=m+n\}P\{\xi=m, \eta=n|\xi+\eta=m+n\}\\
&= \frac{\lambda^{m+n}}{(m+n)!}e^{-\lambda}\cdot\binom{m+n}{n}p^m q^n = \frac{\lambda^m p^m}{m!}e^{-\lambda p}\cdot\frac{\lambda^n q^n}{n!}e^{-\lambda q}\\
&= P\{\xi=m\}\cdot P\{\eta=n\}.
\end{aligned}
$$

由 (6.8) 式知 ξ 与 η 独立.

一般地说，在独立重复的贝努里试验序列中，如果试验重复的次数服从普阿松分布，那么其中成功次数与失败次数相互独立. 这个命题的逆也成立，利用母函数可给出简单的证明. 参见习题 3.3.5 的第 8 题.

2.6.3 习题

1. 在习题 2.5.5 的 1 题中，求已知 $\eta = j$ $(j = 1, 2)$ 时 ξ 的条件分布.

2. 在习题 2.5.5 的 7 题中，求已知 $\xi_2 = n$ 时 $(n = 2, 3, \cdots)$ ξ_1 的条件分布.

3. 在习题 2.5.5 的 3 题中，求已知 $\eta = y \in (0, \pi/2)$ 时 ξ 的条件密度函数.

4. 设随机向量 (ξ, η) 的联合密度

$$p(x, y) = 24y(1-x-y), \qquad x, y > 0 \text{ 且 } x + y < 1,$$

试求 (1) $\xi=1/2$ 条件下 η 的条件密度； (2) $\eta=1/2$ 条件下 ξ 的条件密度.

5. 设随机变量 ξ 有密度函数

$$p_1(x)=\lambda^2 xe^{-\lambda x},\qquad x>0,$$

而 η 服从区间 $(0,\xi)$ 上的均匀分布. 试求 η 的密度函数.

6. 试证常数 c 与任何随机变量 ξ 相互独立.

7. 设随机向量 (ξ,η) 的联合密度函数为

(1) $p(x,y)=4xy,\qquad 0<x,y<1;$

(2) $p(x,y)=8xy,\qquad 0<x<y<1.$

试问 ξ 与 η 是否独立？为什么？

8. 设随机向量 (ξ,η) 的联合密度函数为

$$p(x,y)=\frac{1+xy}{4},\qquad |x|<1\text{ 且 }|y|<1,$$

求证 ξ 与 η 不独立，但 ξ^2 与 η^2 是独立的.

9. 若 ξ,η 独立，都服从 -1 与 1 这两点上的等可能分布，而 $\zeta=\xi\eta$. 求证 ξ,η,ζ 两两独立但整体不独立.

2.7 随机变量函数的分布

2.7.1 问题的提出，离散型

在概率论的实际应用中，时常会遇到求随机变量函数分布的问题. 例如在统计物理中 1, 已知气体分子的速度 ξ 之分布，求分子的动能 $\frac{1}{2}m\xi^2$ 的分布. 又如已知炮弹弹着点的地图坐标 (ξ,η) 的分布，求弹着点到目标 (a,b) 之距离 $\rho=\sqrt{(\xi-a)^2+(\eta-b)^2}$ 的分布等等，都是求随机变量函数分布的例子.

问题的提法十分明确，而且可以立即给出一般解法. 在一元情形，这就是已知随机变量 ξ 的分布函数 $F_\xi(x)$, 求 $\eta=f(\xi)$ 的分布，其中 f 为某个波雷尔函数. 运用 2.1.3 节中 (1.15) 式立得， η 的分

布函数为

$$P\{\eta < y\} = P\{f(\xi) < y\} = \int_{\{x:f(x)<y\}} dF_\xi(x). \qquad (7.1)$$

但是对于具体的问题，计算 (7.1) 式右方的积分往往需要一些技巧. 以下将分几种情形作较详细的讨论.

首先，对于离散型情形，(7.1) 式右方的积分化为求和. 我们以二元情形为例. 设随机向量 (ξ,η) 有 2.5.2 节中 (5.7) 式定义的联合分布 $\{p_{ij}\}$, 而 $\zeta = f(\xi,\eta)$. 那么，对 ζ 的任何可能值 z_k，有

$$P\{\zeta = z_k\} = P\{f(\xi,\eta) = z_k\} = \sum_{(i,j):f(x_i,y_j)=z_k} p_{ij}. \qquad (7.2)$$

因此，只需对 (ξ,η) 的一切可能值 (x_i,y_j) 算出 ζ 的对应值 z_k, 问题便可解决.

对于非负整值随机变量，尚有如下有用的公式.

定理 7.1 离散卷积公式 设 ξ 与 η 是相互独立的非负整值随机变量，各有分布 $\{a_k\}$ 与 $\{b_k\}$. 那么其和有分布

$$P\{\xi+\eta=n\} = \sum_{k=0}^{n} a_k b_{n-k}, \qquad n = 0,1,2,\cdots, \qquad (7.3)$$

证明: 以 $\{\xi = k\}$ 为分割用全概公式，并注意 ξ 与 η 独立便得到，对任何 $n = 0,1,2,\cdots$ 有

$$P\{\xi+\eta=n\} = \sum_{k=0}^{n} P\{\xi=k\}P\{\xi+\eta=n|\xi=k\}$$

$$= \sum_{k=0}^{n} P\{\xi=k\}P\{\eta=n-k\} = \sum_{k=0}^{n} a_k b_{n-k}.$$

定理证完.

例 7.1 设随机变量 ξ 与 η 独立，ξ 服从 $b(k;n,p)$ 分布，η 服从 $b(k;m,p)$ 分布. 求 $\xi+\eta$ 的分布.

解: 将 $b(k;n,p)$ 与 $b(k;m,p)$ 的表达式代入卷积公式 (7.3), 并利用组合性质 (参见 1.2.2 节的定理 2.1) 便得对 $i=0,1,\cdots,n+m$ 有

$$P\{\xi+\eta=i\}=\sum_{k=0}^{i}b(k;n,p)b(i-k;m,p)$$

$$=\sum_{k=0}^{i}\binom{n}{k}\binom{m}{i-k}p^iq^{n+m-i}=\binom{n+m}{i}p^iq^{n+m-i},$$

即 $\xi+\eta$ 服从 $b(i;n+m,p)$ 分布. 这个结果从直观上很容易理解. 将成功率为 p 的试验独立重复 n 次, 取得 ξ 次成功, 再将此试验独立地重复 m 次, 取得 η 次成功. 因此 $\xi+\eta$ 是 $n+m$ 重贝努里试验之成功次数, 应服从参数为 $m+n$ 与 p 的二项分布.

2.7.2 一元连续型

我们将分几种场合分别讨论连续型随机变量函数的分布. 所使用的方法都是从求未知的分布函数入手. 首先考虑一元情形.

例 7.2 设随机变量 ξ 服从区间 $(-\pi/2,\pi/2)$ 上的均匀分布, 求 $\eta=\mathrm{tg}\xi$ 的分布.

解: η 的分布函数为

$$F(y)=P\{\eta<y\}=P\{\mathrm{tg}\xi<y\}=P\{\xi<\mathrm{tg}^{-1}y,\xi\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\}$$

$$=\int_{-\frac{\pi}{2}}^{\mathrm{tg}^{-1}y}\frac{1}{\pi}dx=\frac{1}{\pi}\mathrm{tg}^{-1}y+\frac{1}{2},$$

求导可得 η 有密度函数

$$p(y)=\frac{1}{\pi(1+y^2)}.$$

它是**柯西分布**$C(\lambda,\mu)$ 当 $\lambda=1,\mu=0$ 的特殊情形. 一般柯西分布的密度为

$$p(x)=\frac{1}{\pi}\cdot\frac{\lambda^2}{\lambda^2+(x-\mu)^2}.$$

在本例的推导中，ξ 仅取 $(-\pi/2,\pi/2)$ 中的值起着重要的作用. 只有局限于 $x\in(-\pi/2,\pi/2), y=\mathrm{tg}x$ 是严格上升的连续函数，存在唯一的反函数 $x=\mathrm{tg}^{-1}y$, 此时才有 $\mathrm{tg}x<y$ 等价于 $x<\mathrm{tg}^{-1}y$. 注意在所得到的密度 $p(y)$ 的表达式中，$1/\pi$ 是 ξ 的密度函数，而因子 $1/(1+y^2)$ 恰是反函数 $\mathrm{tg}^{-1}y$ 的导数. 我们叙述如下的一般结果.

定理 7.2 设随机变量 ξ 有密度函数 $p_1(x)$, $x\in(a,b)$, 而 $\eta=f(\xi)$. 如果 $y=f(x)$ 在 $x\in(a,b)$ 上是严格单调的连续函数，存在唯一的反函数 $x=h(y), y\in(\alpha,\beta)$, 并且 $h'(y)$ 存在且连续，那么 η 也是连续型随机变量，具有密度函数

$$p_2(y)=p_1[h(y)]|h'(y)|, \qquad y\in(\alpha,\beta). \tag{7.4}$$

定理的证明与例 7.2 的解题过程类似. 注意当 f 为严格下降函数时，推导中 $f(x)<y$ 等价于 $x>h(y)$, 恰好此时 $h'(y)<0$. 故在 $p_2(y)$ 的表达式 (7.4) 中需取 $h'(y)$ 的绝对值.

最简单的特例是线性函数，即 $f(x)=a+bx$ $(b\neq 0)$. 此时用 (7.4) 式易得，当 ξ 服从 $N(0,1)$ 分布时，$a+b\xi$ 服从 $N(a,b^2)$ 分布；而当 ξ 服从 $U(0,1)$ 分布时，$a+b\xi$ 服从 $U(a,a+b)$ 分布.

例 7.3 设 ξ 服从标准正态分布，求 $\eta=\xi^2$ 的分布.

解: 显然 η 不取负值，因此当 $y\leq 0$ 时，η 的分布函数 $F(y)=0$. 而当 $y>0$ 时

$$F(y)=P\{\xi^2<y\}=P\{-\sqrt{y}<\xi<0\}+P\{0<\xi<\sqrt{y}\}$$

$$=\int_{-\sqrt{y}}^{0}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\,dx+\int_{0}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\,dx.$$

求导可得 η 的密度函数

$$p(y)=\frac{1}{\sqrt{2\pi y}}\,e^{-\frac{y}{2}}, \qquad y>0.$$

这是 n 个自由度的 χ^2 分布密度函数

$$p_n(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\, y^{\frac{n}{2}-1}\, e^{-\frac{y}{2}}, \qquad y > 0 \tag{7.5}$$

当 $n=1$ 时的特例. 以后将证明, 如果 $\xi_1,\cdots,\xi_n$ 相互独立同服从标准正态分布, 则 $\eta = \sum\limits_{k=1}^{n} \xi_k^2$ 服从密度函数由 (7.5) 式给出的 χ^2 分布.

回顾例 7.3 的求解过程. 虽然 ξ 取值于 $x \in (-\infty,+\infty)$, 但是 $y=x^2$ 仅取 $[0,+\infty)$ 中的值, 而且反函数不唯一. 如果分割 $f(x)=x^2$ 的定义域 $(-\infty,+\infty)=(-\infty,0)\cup[0,+\infty)$, 那么在每个子区间上函数 $f(x)$ 是严格单调且连续的, 各有一支反函数 $x=-\sqrt{y}$ 与 $x=\sqrt{y}$. 分别用 (7.4) 式再把它们相加, 便得到本例的密度函数. 我们有如下一般定理.

定理 7.3 设随机变量 ξ 有密度函数 $p_1(x)$, $x\in(a,b)$. 如果可分割 $(a,b)=\bigcup\limits_i I_i$, 而在每个子区间 I_i 上函数 $y=f(x)$ 有唯一的反函数 $h_i(y)$, 且 $h_i'(y)$ 存在连续. 那么 $\eta=f(\xi)$ 仍为连续型变量, 其密度函数

$$p_2(y) = \begin{cases} \sum\limits_i p_1[h_i(y)]\cdot|h_i'(y)|, \\ 0, \qquad\qquad \text{若}y\text{使反函数不存在}. \end{cases} \tag{7.6}$$

定理的证明与例 7.3 的解类似, 不再详述.

2.7.3 多元连续型, 特殊情形

以下考虑随机向量 $\vec{\xi}=(\xi_1,\cdots,\xi_n)$ 的变换 $\vec{\eta}=f(\vec{\xi})$. 本节先对一些常用的特殊 f 导出 $\vec{\eta}$ 的密度函数.

定理 7.4 和的密度 若随机向量 (ξ,η) 有联合密度函数 $p(x,y)$, 则它们的和 $\xi+\eta$ 仍为连续型, 具有密度

$$p_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} p(x,z-x)\,dx = \int_{-\infty}^{+\infty} p(z-y,y)\,dy. \tag{7.7}$$

证明: 还是从 $\xi+\eta$ 的分布函数 $F(z)$ 入手.

$$
\begin{aligned}
F(z) &= P\{\xi+\eta<z\} \\
&= \iint_{\{x+y<z\}} p(x,y)\,dxdy = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} p(x,y)\,dy \\
&= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z} p(x,v-x)dv = \int_{-\infty}^{z}\left[\int_{-\infty}^{-\infty} p(x,v-x)dx\right]dv.
\end{aligned}
$$

至此 $\xi+\eta$ 的分布函数 $F(z)$ 已表为方括号内函数的变上限积分, 于是 $\xi+\eta$ 仍为连续型, 密度如 (7.7) 中第一个表达式所示. 改变积分次序可得 (7.7) 中另一个表达式. 证毕.

特别当 ξ 与 η 独立时, (7.7) 式变为

$$
p_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} p_\xi(x)p_\eta(z-x)\,dx = \int_{-\infty}^{+\infty} p_\xi(z-y)p_\eta(y)\,dy, \quad (7.8)
$$

称这一运算为函数 $p_\xi(x)$ 与 $p_\eta(y)$ 的**卷积**;(7.8) 式即为连续型场合的**卷积公式**. 可以把此式与前面的离散型卷积公式 (7.3) 统一用相应的分布函数写成

$$
F_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} F_\eta(z-x)\,dF_\xi(x) = \int_{-\infty}^{+\infty} F_\xi(z-y)\,dF_\eta(y), \quad (7.9)
$$

通常简写为 $\quad F_{\xi+\eta} = F_\xi * F_\eta$.

例 7.4 设随机变量 ξ,η 独立, 同服从 $\lambda=1$ 的指数分布, 求 $\xi+\eta$ 的密度函数.

解: 注意 $p_\xi(x)=e^{-x}, x>0; p_\eta(z-x)=e^{-(z-x)}, z-x>0$. 故为要使卷积公式中被积函数非 0, 当且仅当 $x>0$ 且 $z-x>0$, 即 $0<x<z$. 当 $z\le 0$ 时这是不可能的, 故此时 $p_{\xi+\eta}(z)=0$; 而当 $z>0$ 时有

$$
p_{\xi+\eta}(z) = \int_0^z e^{-x}\,e^{-(z-x)}\,dx = z\,e^{-z}.
$$

注意指数分布的卷积不再是指数分布. 但是作为伽马分布的特例, ξ 与 η 相互独立同服从 $\Gamma(1,1)$ 分布的随机变量, 其和具有 $\Gamma(1,2)$ 分布. 这正是 $\lambda=1$ 的普阿松过程中第 $r=2$ 个质点到达时间的分布 (参见 2.4.3 节及本节习题 6).

定理 7.5 商的密度 设随机变量 (ξ,η) 有联合密度函数 $p(x,y)$, 则它们的商 ξ/η 仍为连续型, 有密度

$$p_{\xi/\eta}(z)=\int_{-\infty}^{+\infty}|y|p(zy,y)dy=\int_{-\infty}^{+\infty}|x|p(x,zx)dx. \qquad (7.10)$$

证明: 商 ξ/η 的分布函数

$$\begin{aligned}
F(z) &= P\{\xi/\eta<z\}=\iint_{\{x/y<z\}}p(x,y)\,dxdy\\
&=\iint_{\{x<yz,y>0\}}p(x,y)\,dxdy+\iint_{\{x>yz,y<0\}}p(x,y)\,dxdy\\
&=\int_0^{+\infty}dy\int_{-\infty}^{yz}p(x,y)\,dx+\int_{-\infty}^{0}dy\int_{yz}^{+\infty}p(xy,y)\,dx\\
&=\int_0^{+\infty}dy\int_{-\infty}^{z}y\cdot p(yu,y)\,du+\int_{-\infty}^{0}dy\int_{z}^{-\infty}y\cdot p(yu,y)\,du\\
&=\int_{-\infty}^{z}\left[\int_{-\infty}^{+\infty}|y|p(yu,y)\,dy\right]du.
\end{aligned}$$

至此已将 $F(z)$ 写为方括号中函数之变上限积分, 于是 ξ/η 仍为连续型, 密度函数如 (7.10) 中第一表达式所示. 改变积分次序可推出另一表达式. 定理得证.

例 7.5 设随机变量 ξ,η 独立同 $\lambda=1$ 的指数分布, 求商 ξ/η 的密度.

解: (ξ,η) 的联合密度函数为

$$p(x,y)=e^{-x-y},\qquad x>0\text{且 }y>0.$$

为要 (7.10) 式中的被积函数 $p(zy,y)$ 非 0, 必须且只需 $y>0$ 且 $zy>0$. 当 $z\le 0$ 时这是不可能的, 从而 $p_{\xi/\eta}(z)=0$. 而当 $z>0$ 时

有

$$p_{\xi/\eta}(z) = \int_0^{+\infty} y\, e^{-yz-y}\, dy = \frac{1}{(1+z)^2}.$$

定理 7.6 最大最小值的分布 设 $\xi_1,\cdots,\xi_n$ 为相互独立随机变量，它们有共同的分布函数 $F(x)$ 与密度函数 $p(x)$. $\eta_1 = \bigwedge_{k=1}^n \xi_k$ 与 $\eta_2 = \bigvee_{k=1}^n \xi_k$ 分别为它们的最小值与最大值，则 $(\eta_1、\eta_2)$ 有联合密度函数

$$q(x,y) = n(n-1)[F(y)-F(x)]^{n-2}p(x)p(y), \qquad x < y. \qquad (7.11)$$

证明: 我们从 (η_1,η_2) 的联合分布函数 $G(x,y)$ 入手. 注意到 $\{\bigvee_{k=1}^n \xi_k < y\} = \bigcap\limits_{k=1}^n \{\xi_k < y\}$ 而 $\{\bigwedge_{k=1}^n \xi_k \geq x\} = \bigcap\limits_{k=1}^n \{\xi_k \geq x\}$. 于是，当 $x \geq y$ 时有

$$G(x,y) = P\{\eta_1 < x, \eta_2 < y\} = P\{\eta_2 < y\}$$

$$= P\{\bigcap_{k=1}^n [\xi_k < y]\} = [F(y)]^n,$$

而当 $x < y$ 时有

$$G(x,y) = P\{\eta_2 < y\} - P\{\eta_1 \geq x, \eta_2 < y\}$$

$$= P\{\bigcap_{k=1}^n [\xi_k < y]\} - P\{\bigcap_{k=1}^n [x \leq \xi_k < y]\}$$

$$= [F(y)]^n - [F(y) - F(x)]^n.$$

求导便得 (η_1,η_2) 的联合密度 $q(x,y) = \frac{\partial^2}{\partial x \partial y} G(x,y)$ 如 (7.11) 式所示. 定理证毕.

例 7.6 向区间 (0,1) 内独立地任投 n 个点，求这 n 个点全分布在长为 1/2 的子区间内的概率.

解: 对 $k=1,2,\cdots,n$, 以 ξ_k 表示第 k 个点的坐标. 则由题意知 $\xi_1,\cdots,\xi_n$ 独立同 $U(0,1)$ 分布. 所求概率为

$$
\begin{aligned}
&P\left\{\bigvee_{k=1}^{n}\xi_k-\bigwedge_{k=1}^{n}\xi_k<\frac{1}{2}\right\}=\iint_{\{y-x<1/2\}}q(x,y)dxdy\\
&=\iint_{\{0<y-x<\frac{1}{2},0<x,y<1\}}n(n-1)(y-x)^{n-2}dxdy\\
&=\int_0^{\frac{1}{2}}dx\int_x^{x+\frac{1}{2}}n(n-1)(y-x)^{n-2}+\int_{\frac{1}{2}}^{1}dx\int_x^{1}n(n-1)(y-x)^{n-2}dy\\
&=\frac{n}{2^n}+\frac{1}{2^n}=\frac{n+1}{2^n}.
\end{aligned}
$$

2.7.4 多元连续型, 一般情形

求连续型随机向量之函数的分布的最一般的提法是: 设 $(\xi_1,\cdots,\xi_n)$ 有联合密度函数 $p(x_1,\cdots,x_n)$, 对 $j=1,2,\cdots,m$, $\eta_j=f_j(\xi_1,\cdots,\xi_n)$, 求随机向量 $(\eta_1,\cdots,\eta_m)$ 的分布.

首先, 函数 f_j 不能太任意. 每个函数 $f_j(x_1,\cdots,x_n)$ 必须有较好的性质, 以保证 η_j 为随机变量. 如 2.1.1 节中定理 1.3 所述, 当 f_j 是 n 维波雷尔函数, 即对任何 $B\in\mathcal{B}$ 有 $\{(x_1,\cdots,x_n):f_j(x_1,\cdots,x_n)\in B\}\in\mathcal{B}^n$ 时这一要求满足. 其次, 由前面对一元情况的讨论可见, 问题的关键是由方程组 $y_j=f_j(x_1,\cdots,x_n),j=1,\cdots,m$ 中能否解出唯一的连续可微的反函数 $x_i=h_i(y_1,\cdots,y_m),i=1,\cdots,n$. 但是当 $m>n$ 时, 既使对线性函数 f_j, 也不能保证方程组有解. 所以我们只考虑 $m=n$ 情形. $m<n$ 时可以通过补设 $\eta_j=\xi_j,j=m+1,\cdots,n$ 化为 $m=n$ 情形.

下面定理给出如下一般的随机向量变换公式.

定理 7.7 假设连续型随机向量 $(\xi_1,\cdots,\xi_n)$ 有联合密度函数 $p(x_1,\cdots,x_n)$, 而 $\eta_j=f_j(\xi_1,\cdots,\xi_n),j=1,\cdots,n$. 每个 f_j 均为 n 维波雷尔函数. 如果对 $(\eta_1,\cdots,\eta_n)$ 的每一组可能值 $(y_1,\cdots,y_n)$, 方程组

$$
\begin{cases}
y_1=f_1(x_1,\cdots,x_n)\\
\qquad\cdots\qquad\cdots\\
y_n=f_n(x_1,\cdots,x_n)
\end{cases}
\tag{7.12}
$$

有唯一的解

$$\begin{cases} x_1 = h_1(y_1, \cdots, y_n) \\ \quad \cdots \quad \cdots \\ x_n = h_n(y_1, \cdots, y_n) \end{cases}, \tag{7.13}$$

并且每个 $h_k(y_1, \cdots, y_n)$ 有连续的一阶偏导数. 那么随机向量 $(\eta_1, \cdots, \eta_n)$ 是连续型的, 有联合密度函数

$$q(y_1, \cdots, y_n) = \begin{cases} p[h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)] \cdot |J|, \\ 0, \qquad \text{若 } (y_1, \cdots, y_n) \text{使 (7.11) 无解,} \end{cases} \tag{7.14}$$

其中 J 为变换 (7.13) 的雅可比行列式.

证明: $(\eta_1, \cdots, \eta_n)$ 的联合分布函数

$$\begin{aligned} F(z_1, \cdots, z_n) &= P\{\eta_1 < z_1, \cdots, \eta_n < z_n\} \\ &= \int_{f_n(x_1,\cdots,x_n)<z_n} \cdots \int_{f_1(x_1,\cdots,x_n)<z_1} p(x_1, \cdots, x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_n} p[h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)] \, |J| dy_1 \cdots dy_n. \end{aligned}$$

最后一步运用了积分的变数替换 (7.13), 于是得证 $(\eta_1, \cdots, \eta_n)$ 的联合密度函数如 (7.14) 式所示.

例 7.7 设坐标平面上随机点的直角坐标 (ξ, η) 相互独立, 同服从 $N(0,1)$ 分布. 求其极坐标 (ρ, θ) 的分布.

解: 本题所涉及的变换

$$\begin{cases} x = r \cos t \\ y = r \sin t \end{cases}$$

是 (ξ, η) 的值域 $\Re^2$ 到 (ρ, θ) 的值域 $[0, +\infty) \times [0, 2\pi)$ 之间的一一变换 (坐标原点除外), 其雅可比行列式 $J = r$. 现知 (ξ, η) 的联合密度函数为

$$p(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)},$$

用公式 (7.14) 便得 (ρ,θ) 有联合密度函数

$$q(r,t)=\frac{1}{2\pi}re^{-\frac{r^2}{2}},\qquad r>0,\ \ t\in[0,2\pi).$$

至此可得 (ρ,θ) 相互独立， θ 服从 $U(0,2\pi)$ 分布而 ρ 的边缘密度函数为

$$q_1(r)=re^{-\frac{r^2}{2}},\qquad r>0,$$

称之为**瑞利分布**.

例 7.8 设随机变量 (ξ,η) 有联合密度函数

$$p(x,y)=3x,\qquad 0<y<x<1.$$

试求 $\xi-\eta$ 的分布.

解: 为求 $\alpha=\xi-\eta$ 的分布，补设 $\beta=\eta$. 于是它们的可能值之间相应地有一一对应

$$\begin{cases}u=x-y\\ v=y\end{cases},\qquad 0<y<x<1.$$

解出

$$\begin{cases}x=u+v\\ y=v\end{cases},\qquad u,v>0\ \text{且}\ u+v<1.$$

此变换的雅可比行列式 $J=1$. 故用 (7.14) 式可得 $(\xi-\eta,\eta)$ 有联合密度函数

$$q(u,v)=3(u+v),\qquad u,v>0\ \text{且}\ u+v<1.$$

注意，此例中 α 与 β 不独立. 积分可得 $\alpha=\xi-\eta$ 的边缘密度函数，当 $0<u<1$ 时

$$q_1(u)=\int_{-\infty}^{+\infty}q(u,v)dv=\int_0^{1-u}3(u+v)dv=\frac{3}{2}(1-u^2),$$

对于区间 (0,1) 之外的 u

$$q_1(u)=\int_{-\infty}^{+\infty}q(u,v)dv=\int_{-\infty}^{+\infty}0dv=0.$$

由此例可见，在运用 (7.14) 式算随机向量函数的分布时，正确地确定变换后随机向量的取值范围是非常重要的.

定理 7.2 中 (7.4) 只是 (7.14) 式当 $n=1$ 时的特例. 可以建立定理 7.3 在 n 维场合的一般结果. 就是说，当方程组 (7.12) 的解不唯一时，如果可将 $(\xi_1,\cdots,\xi_n)$ 的可能值集 G 分割为不交并 $G=\bigcup\limits_i G_i$，使得在每个 G_i 上 (7.12) 都有满足定理中条件的唯一解，那么，我们可在每个 G_i 上运用 (7.14) 式，然后对一切可能的 i 求和便得到类似 (7.6) 的最一般公式. 在此不加详述.

2.7.5 随机变量的存在性

在概率论中，我们经常在 "设随机变量 ξ 的分布函数为 $F(x)$" 的假定下讨论问题. 那么，这样的假定是否合理呢？回答当然是肯定的. 给定 $\Re$ 上单调非降, 左连续, 有下确界 0 与上确界 1 的函数 $F(x)$, 总存在概率空间及其上的随机变量 ξ, 使 ξ 以 $F(x)$ 为分布函数.

本章 2.1.3 节介绍随机变量的概率分布时曾讲过. 每个满足上述条件的分布函数 $F(x)$, 总可以在 $\Re$ 上的波雷尔函数类 $\mathcal{B}$ 上引进一个勒贝格 – 斯蒂尔吉斯测度 $\mathbf{F}$, 使得

$$\mathbf{F}\{(-\infty,x)\}=F(x),\qquad \text{任}\ \ x\in\Re.$$

于是可将概率空间取为 $(\Re,\mathcal{B},\mathbf{F})$. 再令 $\xi(\omega)=\omega$, 任 $\omega\in\Re$. 那么 ξ 的分布函数为

$$P\{\xi(\omega)<x\}=P\{\omega<x\}=\mathbf{F}\{(-\infty,x)\}=F(x).$$

满足我们的要求. 不过，上述论证的严格性是建立在测度扩张定理之上的. 我们没有给出证明. 以下是随机变量存在性定理的另一个构造性的证明.

熟知每个 $\Re$ 上严格单调的连续函数 $F(x)$ 都有唯一的反函数 $F^{-1}(x)$; 反函数与 $F(x)$ 有相同的单调性与连续性. 现分布函数只有非降与左连续性，需要将反函数的概念加以推广.

定义 7.1 称

$$F^{-1}(y) = \inf\{x : F(x) > y\}, \qquad y \in (0,1), \qquad (7.15)$$

为分布函数 $F(x)$ 的**单调逆**.

易见， $F^{-1}(y)$ 是单调非降且右连续的，而且当 $F(x)$ 严格上升的连续函数时，上述单调逆就是普通的反函数.

图 7.1 画出了分布函数 $y = F(x)$ 及其单调逆 $x = F^{-1}(y)$ 的图形.

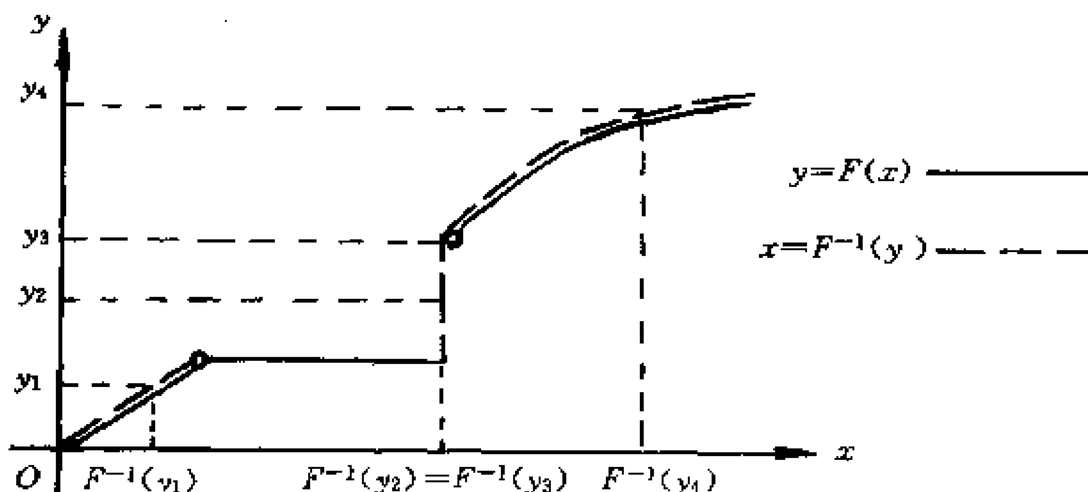


图 7.1 函数 $F(x)$ 与 $F^{-1}(x)$ 的图形

定理 7.8 设 $F(x)$ 为 $\mathfrak{R}$ 上的分布函数， ξ 为区间 (0,1) 上的均匀分布的随机变量，则随机变量 $\eta = F^{-1}(\xi)$ 以 $F(x)$ 为分布函数.

证明: 由单调逆 F^{-1} 的定义 (7.15) 式及 $F(x)$ 的左连续性可知，

$$F(x) > y \quad \text{当且仅当} \quad x > F^{-1}(y), \qquad (7.16)$$

从而对一切实数 x 有

$$P\{\eta < x\} = P\{F^{-1}(\xi) < x\} = P\{\xi < F(x)\} = F(x).$$

这表明 η 的分布函数是 $F(x)$. 证毕.

定理 7.9 对任何给定的分布函数 $F(x)$, 都存在概率空间 $(\Omega,\mathcal{F},P)$ 及其上的随机变量 η, 使 η 以 $F(x)$ 为分布函数.

证明: 取 $\Omega=(0,1),\mathcal{F}$ 为区间 (0,1) 上的波雷尔集类, P 为 Ω 上的勒贝格测度. 则 $(\Omega,\mathcal{F},P)$ 为一概率空间. 再取 $\xi(\omega)=\omega$, 则 ξ 为区间 (0,1) 上的均匀分布的随机变量. 据定理 7.8 知, $\eta=F^{-1}(\xi)$ 即为所求以 $F(x)$ 为分布函数的随机变量. 证完.

随机变量存在性的这个构造性证明方法, 与本节开头所述的方法有所不同. 那里的概率空间是取作由 $F(x)$ 引出的勒贝格 — 斯蒂尔吉斯测度空间 $(\Re,\mathcal{B},\mathbf{F})$, 即概率空间随 $F(x)$ 不同而不同. 而这里总取区间 (0,1) 上的勒贝格测度空间, 与 $F(x)$ 无关. 因此可在其上构造出具有不同分布的多个随机变量. 这一点今后会用到.

2.7.6 随机数

在概率论的应用中, 常常需要模拟具有指定分布的随机变量. 就是说需要给出某种随机变量的具体试验结果 (即样本值). 例如, 从仓库中摆放着的 n 台电视机中任取一台进行检验, 要实现这个“任取”, 就要模拟 n 个点上等可能分布的随机变量 (可借助抽签实现). 又如在 1.3.1 节中讨论几何概型时, 我们曾介绍过蒙特卡洛方法的一个实例, 那就是通过蒲丰投针试验来计算圆周率 π 的近似值. 但是, 真要成千上万次地做投针试验, 那是相当费时费力的. 可以采用随机模拟的方法. 在这个例子中, 我们只需要统计出事件 $\{(\rho,\theta):\rho<\frac{l}{2}\sin\theta\}$ 发生的频率, 其中刻画针的位置的 ρ 与 θ 独立, 分别在区间 $(0,a/2)$ 与 $(0,\pi/2)$ 上均匀分布. 于是, 每当我们给出一对按上述分布的数值 (ρ,θ) 时, 就相当完成一次投针试验.

服从区间 (0,1) 上均匀分布的随机变量的值, 称为**均匀分布随机数**, 简称**随机数**. 它是产生其它分布随机数的基础. 至于均匀分布随机数的产生, 从原理上讲, 可以将 10 张外形相同的卡片上分别写上数字 0 至 9, 然后有放回地一一取出卡片. 这样, 各次取出卡片上的数字是相互独立的, 且在 $0,1,\cdots,9$ 这 10 个数字上等可能分布. 将

依次取出的数字 $a_1,\cdots,a_n$ 前加上小数点，于是 $0.a_1a_2\cdots a_n$ 就是从区间 (0,1) 上任取（均匀分布）的一个数. 多取几位可以获得更大的精度. 应用中常常设计一定的算法，使算得结果经检验后可当作随机数使用，称作**伪随机数**. 将它们列成随机数表备查（见本书附表 IV），或将计算方法编制成软件存放计算机中，随时可输出随机数备用.

下面介绍运用均匀分布随机数产生其它分布的随机数的方法. 从数学上讲，就是已知随机变量 ξ 服从 $U(0,1)$ 分布，寻求适当的函数 f，使得 $f(\xi)$ 有指定的分布函数 $F(x)$. 据定理 7.8 知，取 $f=F^{-1}$ 即可.

下面介绍具体做法，总设 ξ 有区间 (0,1) 上的均匀分布.

先看离散型情形. 设法构造有指定的离散型分布 $P\{\eta=y_j\}=p_j, j=1,2,\cdots$ 的随机变量 η. 将区间 (0,1) 分划成相邻的小区间 $I_j=(p_1+\cdots+p_{j-1}, p_1+\cdots+p_j)(j=1,2,\cdots)$ 之并（取 $p_0=0$），显然每个 I_j 的长恰为 p_j. 再令

$$\eta=\sum_j y_j\chi_{I_j}(\xi),$$

这里 $\chi_{I_j}(\xi)$ 是区间 I_j 的示性函数. 于是有

$$P\{\eta=y_j\}=P\{\xi\in I_j\}=p_j, \qquad j=1,2,\cdots.$$

这说明所构造的 η 正是所要的离散型随机变量.

例 7.9 试构造具有分布

$$P\{\eta=j\}=\frac{j}{10}, \qquad j=1,\ 2,\ 3,\ 4$$

的随机数五个.

解: 显然应取 $I_1=(0,0.1)$, $I_2=(0.1,0.3)$, $I_3=(0.3,0.6)$, $I_4=(0.6,\ 1.0)$. 查附表 IV 得前五个随机数 0.04839, 0.96423, 0.24878, 0.82651, 0.66566. 易见它们依次在 I_1, I_4, I_2, I_4, I_4 中，故得符合要求的五个随机数为 1, 4, 2, 4, 4.

以下给出产生连续型随机数的例子.

例 7.10 构造五个服从 $\lambda=1$ 指数分布的随机数.

解: $\lambda=1$ 指数分布的分布函数为

$$F(x)=1-e^{-x}, \qquad x\in(0,+\infty).$$

解出其反函数

$$F^{-1}(y)=-\ln(1-y), \qquad y\in(0,1).$$

于是 $-\ln(1-\xi)$ 就是 $\lambda=1$ 指数分布随机变量，由于 $1-\xi$ 与 ξ 同为 $U(0,1)$ 分布，故可采用 $-\ln\xi$, 再将例 7.9 中引用的五个均匀分布随机数代入可得所要的 $\lambda=1$ 指数分布随机数. 它们是 3.02846, 0.03643, 1.39119, 0.19054, 0.40698.

例 7.11 正态分布随机数的产生 第一种方法是直接运用定理 7.8. 虽然正态分布的分布函数不是初等函数，无法解出显式的反函数 $\Phi^{-1}(y)$. 但是可通过反查标准正态分布函数的数值表 (附表 III) 而得到标准正态随机数. 比如例 7.9 中引用的五个随机数所对应的标准正态随机数依次为 -1.66, 1.80, -0.68, 0.94, 0.43. 再经线性变换便可得到任何参数的正态分布随机数.

下面介绍另一种方法. 据例 7.7 可知，如果 ρ 与 θ 独立，ρ 服从瑞利分布，θ 服从 $U(0,2\pi)$ 分布，那么

$$\eta_1=\rho\cos\theta \qquad \text{与} \qquad \eta_2=\rho\sin\theta$$

是一对独立的 $N(0,1)$ 随机变量. 现 $U(0,2\pi)$ 的随机数易由 $U(0,1)$ 随机数得到，而瑞利分布的分布函数为 $F(x)=1-e^{-\frac{x^2}{2}}$, 解出反函数

$$F^{-1}(y)=[-2\ln(1-y)]^{1/2}.$$

至此，运用定理 7.7 可知，如果 ξ_1,ξ_2 为二独立的均匀分布随机数，则由

$$\eta_1=[-2\ln\xi_1]^{1/2}\cos(2\pi\xi_2)$$

与

$$\eta_2 = [-2\ln\xi_1]^{1/2}\sin(2\pi\xi_2)$$

给出两个相互独立的标准正态随机数.

2.7.7 习题

1. 设 ξ,η 独立，同服从参数为 p 的几何分布，试求 (1) $\xi+\eta$; (2) $\xi\vee\eta$ 的分布.

2. 设随机向量 (ξ,η) 有联合分布如下表.

p_{ij}		y_j				$p_{i\cdot}$
		-1	0	1	2	
x_i	1	2/16	0	2/16	1/16	5/16
	2	0	3/16	0	2/16	5/16
	3	2/16	0	1/16	3/16	3/16
$p_{\cdot j}$		4/16	3/16	3/16	6/16	1

试求 (1)$\xi+\eta$ (2)$\xi\bigvee\eta$ (3)η^2 的概率分布.

3. 设随机变量 ξ_1,ξ_2 相互独立，对 $i=1,2,\xi_i$ 服从参数为 λ_i 的普阿松分布，试求 (1)$\xi_1+\xi_2$ 的分布; (2) 已知 $\xi_1+\xi_2=n$ 时 ξ_1 的条件分布.

4. 设随机变量 ξ 有标准正态分布，求 (1)e^{ξ}; (2)$1/\xi^2$ 的密度函数.

5. 设 $\xi_1,\cdots,\xi_n$ 为相互独立的随机变量，每个 ξ_i 服从参数为 λ_i 的指数分布. 试求它们的最小值 $\bigwedge_{i=1}^n\xi_i$ 的分布.

6. 设随机变量 ξ 有密度函数 $p(x)$, 试求 (1) ξ^{-1} (2) $\text{tg}\xi$ (3)$|\xi|$ 的分布函数.

7. 设随机变量 ξ 与 η 独立，同服从 $U(0,1)$ 分布，求 $\xi+\eta$ 的密度函数.

8. 设随机变量 $\xi_1,\cdots,\xi_r$ 相互独立，都服从参数为 λ 的指数分布. 试证 $\xi_1+\cdots+\xi_r$ 服从伽马分布 $\Gamma(\lambda,r)$.

9. 设 ξ, η 为独立同 $U(0,a)$ 分布 $(a>0)$ 的随机变量，试求 ξ/η 的密度函数.

10. 向区间 $(0,a)$ 内任投两点，求两点间距离的密度函数.

11. 设 (ξ,η) 服从区域 $D=\{(x,y): 0<|x|<y<1\}$ 上的均匀分布，试求 ξ^2 的密度函数.

12. 设随机向量 (ξ,η) 的联合密度函数为

$$p(x,y)=\frac{1}{x^2y^2}, \qquad x,y\geq 1.$$

试求 $\alpha=\xi\eta$ 与 $\beta=\xi/\eta$ 的联合密度函数与边缘密度函数.

13. 设随机变量 ξ,η 独立，同服从标准正态分布. 试求 $\xi+\eta$, $\xi-\eta$ 的联合分布与边缘分布.

14. 试证上题中 ξ/η 服从柯西分布.

15. 设随机变量 ξ,η 独立同 $\lambda=1$ 的指数分布，试证 $\xi+\eta$ 与 $\xi/(\xi+\eta)$ 相互独立.

16. 设随机变量 ξ 与 η 相互独立，同服从 $\Gamma(\lambda,r)$ 分布. 试证 $\xi+\eta$ 与 ξ/η 相互独立.

17. 设平面上随机点的直角坐标 (ξ,η) 有联合密度函数

$$p(x,y)=\frac{2}{\pi}(1-x^2-y^2), \qquad 0<x^2+y^2<1.$$

求此点极坐标 (ρ,θ) 的联合密度与边缘密度函数.

18. 设随机向量 (ξ,η,ζ) 有联合密度函数

$$p(x,y,z)=6(1+x+y+z)^{-4}, \qquad x,y,z>0.$$

试求 $\xi+\eta+\zeta$ 的密度函数.

19. 设随机变量 ξ 与 η 独立. ξ 服从 $U(0,1)$ 分布， η 的分布函数为

$$F(y)=1-\frac{1}{y^2}, \qquad y>1.$$

试求 $\xi\eta$ 的密度函数.

20. 设 ξ,η 为独立随机变量，ξ 服从 $p=1/2$ 的贝努里分布，η 服从区间 (0,1) 上的均匀分布. 试求

$$(1)\xi+\eta; \qquad (2)\frac{\xi}{2}+\eta; \qquad (3)\xi\eta$$

的分布函数.

21. 设 ξ_1,ξ_2 独立，对 $i=1,2$，ξ_i 服从区间 $(0,i)$ 上的均匀分布. 试用随机数表计算 $P\{\xi_1+\xi_2<1\}$ 的近似值，并与其精确值相比较.

22. 设 ξ,η 独立同服从 $N(0,1)$ 分布. 试用随机数表计算 $P\{\xi\eta>1\}$ 的近似值.

第三章　数字特征与特征函数

3.1　数学期望

3.1.1　初等定义

概率分布完全描述了随机试验中随机变量的概率性质．然而在一些实际问题中，人们并不需要知道随机变量的全部概率性质，而是只需从某个侧面了解其部分概率性质．例如一批棉花中棉纤维的长度，强度等都有客观存在的概率分布．但是当我们用棉花纺纱织布时，只需掌握棉纤维长度，强度的平均值，便可确定其对棉布质量的影响．

象平均值这种表现概率分布的某种特征之数量，就是概率分布的数字特征．

由于数字特征更集中地反映了概率分布在某方面的性质，故在一定的场合中它更便于应用．而且在常见的概率分布中，所含的参数往往都是此分布的数字特征．于是，在已知分布的类型之后，这种分布由其数字特征完全确定．引进数字特征的另一原因是，有时很难精确地求出概率分布，因之不得已退而求其次．本节介绍数学期望，它是最基本的数字特征，先看一个初等的例子．

例 1.1　在一次考试中，10 名学生有 2 人得 70 分，5 人得 80 分，3 人得 90 分，那么他们的平均成绩为 81 分．具体算法是

$$(70 \times 2 + 80 \times 5 + 90 \times 3) \div 10 = 70 \times 0.2 + 80 \times 0.5 + 90 \times 0.3.$$

换一个观点看，从这 10 个学生中任取 1 人，其考试成绩是随机变量 ξ，有分布

$$\begin{pmatrix} 70 & 80 & 90 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}$$

上述平均分算式的右端正好是 ξ 的各可能值与相应概率乘积之和. 一般地说, 离散型随机变量的平均值应为 $\sum_i x_i P\{\xi = x_i\} = \sum_i x_i p_i$.

我们所以举出这个简单的例子, 是想在一开头就向读者明确地说明, 由 $\sum_i x_i P\{\xi = x_i\}$ 所定义的数字特征, 确确实实是随机变量 ξ 的平均值.

考虑到 ξ 会有无穷多个可能值 x_i, 同时这些 x_i 可正可负, 而平均值应当与求和的次序无关. 反映在数学上便是要求级数 $\sum_i x_i p_i$ 绝对收敛. 对于连续型情形, 以密度函数 $p(x)$ 代替 $\{p_i\}$, 并相应地以积分代替求和, 于是我们有如下的初等定义.

定义 1.1 设 ξ 为离散型随机变量, 有概率分布 $p_i = P\{\xi = x_i\}$, $i = 1, 2, \cdots$. 如果 $\sum_i |x_i| p_i < +\infty$, 则称

$$E(\xi) = \sum_i x_i p_i \tag{1.1}$$

为 ξ (或其分布) 的**数学期望**. 对具有密度函数 $p(x)$ 的连续型随机变量 ξ, 如果 $\int_{-\infty}^{+\infty} |x| p(x)\, dx < +\infty$, 则称

$$E(\xi) = \int_{-\infty}^{+\infty} x p(x)\, dx \tag{1.2}$$

为 ξ (或其分布) 的**数学期望**. 数学期望简称**期望**, 又称**均值**. 如果上述绝对收敛条件不满足, 则称随机变量 (或其分布) 的数学期望不存在.

可将 (1.1) 与 (1.2) 式统一写作

$$E(\xi) = \int_{-\infty}^{+\infty} x\, dF(x), \tag{1.3}$$

其中 $F(x)$ 为 ξ 的分布函数. 如前所述, (1.3) 式右方实际上是勒贝格 - 斯蒂尔吉斯积分, 它在离散型与连续型场合的具体化就是 (1.1) 与 (1.2) 式.

从以下例题可以看到 "期望" 一词的历史渊源.

例 1.2 十万张奖券为一组，每组设头奖2名各得奖金10000 元，二等奖 20 名各得奖金 1000 元，三等奖 200 名各得奖金 100 元，四等奖 2000 名各得奖金 10 元，五等奖 10000 名各得奖金 2 元. 于是买一张奖券能获得的奖金数 ξ 服从如下分布

$$\begin{pmatrix} 10000 & 1000 & 100 & 10 & 2 & 0 \\ 2\times 10^{-5} & 2\times 10^{-4} & 2\times 10^{-3} & 2\times 10^{-2} & 10^{-1} & 0.87778 \end{pmatrix}$$

注意，如果忘掉“末等奖 0 元”而将矩阵的最后一列删去，那么就不成为概率分布了. 对于一个买了一张奖券等待开奖的人来说，他“期望”的奖金数就是 ξ 的平均值 $E(\xi)=0.2+0.2+0.2+0.2+0.2+0=1$(元). 就是说，他必须再去买一张奖券，才能期望开奖时得到一个 2 元的五等奖.

如下例子是数学期望的一个简单应用.

例 1.3 节约化验费的方案 在一个有多名职工的单位中验血普查肝炎病毒. 若每个人的血样分别化验，则一个人耗费一份化验费. 现采用以下改进方案：先将每个人的血样各取出一部分，k 个人为一组混合后化验. 如果呈阴性，则 k 个人同时通过，每人化验 $1/k$ 次. 如果呈阳性，再将 k 个人的血样分别化验，以找出血中含病毒者. 这样每个人化验 $(1+1/k)$ 次. 假定职工中血液不含肝炎病毒的概率为 q，且各职工的情况相互独立. 那么采用改进方案，每位职工化验次数 ξ 的分布为

$$P\left\{\xi=\frac{1}{k}\right\} = 1-P\left\{\xi=1+\frac{1}{k}\right\} = q^k.$$

从而每位职工的平均化验次数

$$E(\xi)=\frac{1}{k}q^k+\left(1+\frac{1}{k}\right)(1-q^k)=1+\frac{1}{k}-q^k.$$

对于给定的 q，选择每组人数 k，只要 $1/k-q^k<0$，则 $E(\xi)<1$，节约了化验费用. 同时，还可选择适当的 k，使平均化验费达到最小.

3.1.2 对概率测度的积分 *

我们已经在 1.4.3 节及 2.1.1 节分别给出概率空间及其上随机变量的严格数学定义. 概率空间 $(\Omega, \mathcal{F}, P)$ 就是一个规范的 $(P\{\Omega\}=1)$ 测度空间, 而随机变量 ξ 就是 $(\Omega, \mathcal{F}, P)$ 上的实值可测函数. 我们还在 2.1.2 节介绍了随机变量的结构. 每个随机变量均可写为两个非负随机变量之差

$$\xi=\xi^{+}-\xi^{-}, \quad \text{其中}\ \xi^{+}=\xi \vee 0, \quad \xi^{-}=(-\xi) \vee 0. \tag{1.4}$$

而非负随机变量 ξ 均可写为非降的非负简单随机变量序列 $\{\xi_n\}$ 的极限. 即存在简单随机变量列 $\{\xi_n\}$, 使得

$$0 \leq \xi_1(\omega) \leq \xi_2(\omega) \leq \cdots \leq \xi_n(\omega) \uparrow \xi(\omega), \qquad \omega \in \Omega. \tag{1.5}$$

最后, 简单随机变量 ξ 就是 $\mathcal{F}$ 中有限个互不相容事件之示性函数的线性组合

$$\xi(\omega)=\sum_{i=1}^{n} a_i \chi_{A_i}(\omega), \qquad A_1, \cdots, A_n \in \mathcal{F}, \tag{1.6}$$

其中

$$\chi_A(\omega)=\begin{cases}1, & \text{若}\ \omega \in A, \\ 0, & \text{若}\ \omega \notin A.\end{cases} \tag{1.7}$$

下面来定义 $(\Omega, \mathcal{F}, P)$ 中随机变量 ξ 对概率测度 P 的积分. 当 ξ 为 (1.7) 式给出的示性函数时, 自然取

$$\int_{\Omega} \xi(\omega) P(d\omega)=P(A). \tag{1.8}$$

再由线性定义 (1.6) 式给出的简单函数的积分

$$\int_{\Omega} \xi(\omega) P(d\omega)=\sum_{i=1}^{n} a_i P(A_i). \tag{1.9}$$

对非负随机变量 ξ, 满足 (1.5) 式的各非负简单函数 ξ_n 的积分是单调非降的，以其极限定义 $\xi(\omega)$ 的积分

$$\int_{\Omega}\xi(\omega)P(d\omega)=\lim_{n}\int_{\Omega}\xi_n(\omega)P(d\omega). \qquad (1.10)$$

当上述极限有限时，称 ξ 为可积的. 最后，如果 (1.4) 式中的 ξ^+ 与 ξ^- 均可积，则称随机变量 ξ 可积，定义其积分

$$\int_{\Omega}\xi(\omega)P(d\omega)=\int_{\Omega}\xi^+(\omega)P(d\omega)-\int_{\Omega}\xi^-(\omega)P(d\omega). \qquad (1.11)$$

可以证明，这样定义的积分是一义存在的，比如非负随机变量的积分 (1.10) 与满足 (1.5) 式的 $\{\xi_n\}$ 的取法无关.

如果去掉概率测度的规范性 $[P(\Omega)=1]$ 要求，而将 P 换为任意其它测度 μ, 类似可定义 Ω 上可测函数 $\xi(\omega)$ 的积分 $\int_{\Omega}\xi(\omega)\mu(d\omega)$. 取 Ω 为 $\Re$, μ 为其上的勒贝格测度 (即长度), 就得到了实变量可测函数的勒贝格积分. 在此不一一详述. 回到随机变量对概率测度的积分，就是数学期望.

定义 1.2 设 $\xi(\omega)$ 是概率空间 $(\Omega,\mathcal{F},P)$ 上的可积随机变量，称

$$E(\xi)=\int_{\Omega}\xi(\omega)P(d\omega) \qquad (1.12)$$

为 ξ 的**数学期望**.

由 (1.8) 式知，事件的概率就是其示性函数 χ_A 的数学期望.

至此，我们给出了数学期望的两种定义，它们之间有什么关系呢？回忆在 2.1.3 节中介绍的随机变量 ξ 的概率分布 $\mathbf{F}$, 它是定义于 $(\Omega,\mathcal{F},P)$ 取值于 $(\Re,\mathcal{B})$ 的随机变量 ξ 在 $\mathcal{B}$ 上引出的概率测度

$$\mathbf{F}(B)=P\{\xi\in B\}, \qquad B\in\mathcal{B}. \qquad (1.13)$$

注意 $(\Re,\mathcal{B},\mathbf{F})$ 也是一个概率空间，其上随机变量 (即波雷尔函数) 的积分与刚才建立的对测度 P 的积分含义相同. 如下著名定理揭示了数学期望的两种定义之间的关系，我们只叙述此定理的一个特殊形式.

定理 1.1 积分变换定理 设 ξ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，$\mathbb{F}$ 为 ξ 通过 (1.13) 式在 $\mathcal{B}$ 上引出的概率测度，则对任何波雷尔函数 $f(x)$ 有

$$\int_{\Re} f(x)\mathbb{F}(dx) = \int_{\Omega} f[\xi(\omega)]P(d\omega). \tag{1.14}$$

上式在如下意义下成立：如果等式一端积分有限，则另一端也有限，且二者相等.

我们指出，当 $f(x)$ 为某波雷尔集 B 的示性函数时，由积分定义及 (1.13) 式知 (1.14) 式为真. 然后逐步按积分定义由简到繁地讨论简单函数，非负波雷尔函数到一般波雷尔函数，便可得证 (1.14) 式正确，在此不加详述.

在 (1.14) 式中取 $f(x) = x$，则等式的右端正是数学期望的定义 (1.12) 式. 而关于概率分布 $\mathbb{F}$ 的积分就是关于 ξ 的分布函数 $F(x)$ 的勒贝格 - 斯蒂尔吉斯积分，故此时 (1.14) 式的左端正好是数学期望初等定义的 (1.3) 式. 定理 1.1 保证两个定义是一致的. 在通常的数学期望初等定义中，不涉及随机变量的样本概率空间 $(\Omega, \mathcal{F}, P)$，只采用它在相空间 $(\Re, \mathcal{B}, \mathbb{F})$ 上的表现形式.

我们顺便运用期望的定义 (1.12) 式证明有关数学期望的一个重要结果.

定理 1.2 设 ξ, η 为概率空间 $(\Omega, \mathcal{F}, P)$ 上两个相互独立的随机变量，且 ξ 与 η 均可积. 那么乘积 $\xi\eta$ 也可积，且

$$E(\xi\eta) = E(\xi) \cdot E(\eta). \tag{1.15}$$

证明: 先假定 ξ, η 为非负简单随机变量

$$\xi = \sum_{i=1}^{n} a_i \chi_{A_i}, \qquad \eta = \sum_{j=1}^{m} b_j \chi_{B_j},$$

其中每个 A_i 与 B_j 为相互独立随机事件．那么它们的乘积为 $\xi\eta=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}a_ib_j\chi_{A_iB_j}$．故用期望定义知

$$E(\xi\eta)=\sum_{i=1}^{n}\sum_{j=1}^{m}a_ib_jP(A_iB_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}a_ib_jP(A_i)P(B_j)$$

$$=\sum_{i=1}^{n}a_iP(A_i)\cdot\sum_{j=1}^{m}b_jP(B_j)=E(\xi)E(\eta).$$

即 (1.15) 式对非负简单随机变量成立．再设 ξ,η 为独立的非负随机变量，按 (1.5) 式取非负简单随机变量列 $\{\xi_n\}$ 与 $\{\eta_n\}$，则 $\{\xi_n\eta_n\}$ 是对 $\xi\eta$ 而言是满足 (1.5) 式的简单随机变量列．由于 ξ 与 η 独立，故 ξ_n 与 η_n 也相互独立，从而有

$$E(\xi\eta)=\lim_{n}E(\xi_n\eta_n)=\lim_{n}E(\xi_n)E(\eta_n)$$

$$=\lim_{n}E(\xi_n)\cdot\lim_{n}E(\eta_n)=E(\xi)E(\eta),$$

并且 $E(\xi)$ 与 $E(\eta)$ 有限保证了 $E(\xi\eta)$ 有限．最后对一般的 ξ,η，由于 ξ^+,ξ^- 与 η^+,η^- 相互独立，且 $E(\xi^+),E(\xi^-),E(\eta^+),E(\eta^-)$ 均有限．因为非负情形 (1.15) 式已真，故有

$$E(\xi\eta)=E[(\xi^+-\xi^-)(\eta^+-\eta^-)]$$

$$=E(\xi^+)E(\eta^+)-E(\xi^+)E(\eta^-)-E(\xi^-)E(\eta^+)+E(\xi^-)E(\eta^-)$$

$$=E(\xi)E(\eta).$$

定理证完．

实际上，我们还有更一般的结果，即通过乘积期望的性质，给出两随机变量相互独立的充分必要条件．

定理 1.3 概率空间 $(\Omega,\mathcal{F},P)$ 上两随机变量 ξ 与 η 相互独立的充分必要条件是，对于使 $f(\xi)$ 与 $g(\eta)$ 均可积的任何波雷尔函数 f,g 有

$$E\left[f(\xi)g(\eta)\right]=E\left[f(x)\right]E\left[g(\eta)\right].\tag{1.16}$$

证明: 必要性 由 2.6.2 节的定理 6.5 知 $f(\xi)$ 与 $g(\eta)$ 仍独立，再运用定理 1.2 便得 (1.16) 式.

充分性 对任何波雷尔集 B_1 与 B_2, 取

$$f(x)=\chi_{B_1}(x), \qquad g(x)=\chi_{B_2}(x).$$

运用 (1.16) 可推出

$$P\{\xi\in B_1,\eta\in B_2\}=E\left[\chi_{B_1}(\xi)\chi_{B_2}(\eta)\right]$$

$$=E\left[\chi_{B_1}(\xi)\right]E\left[\chi_{B_2}(\eta)\right]=P\{\xi\in B_1\}\,P\{\eta\in B_2\},$$

此即 ξ 与 η 相互独立. 定理得证.

3.1.3 数学期望的性质

我们继续讨论由 (1.1),(1.2) 或 (1.3) 式所定义的数学期望. 如果 (ξ,η) 是联合分布着的随机向量，比如有联合密度函数 $p(x,y)$, 那么对于 ξ 而言，其期望

$$E(\xi)=\int_{-\infty}^{+\infty}xp_1(x)\,dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)\,dxdy.$$

这就是说，分量 ξ 的期望既可以用它的边缘分布来求，又可以用联合分布直接按最后一个积分求出. 离散型情形此结论仍然成立. 最后，随机向量的期望定义为各分量之期望组成的向量.

作为积分，数学期望具备积分的通常性质.

定理 1.4 假设所涉及的随机变量都有有限的数学期望，我们有

1° $E(c)=c$;

2° $E(c\xi)=cE(\xi)$;

3° $E(\xi+\eta)=E(\xi)+E(\eta)$;

4° 若 $\xi\geq 0$, 则 $E(\xi)\geq 0$;

5° 若 $\xi\geq\eta$, 则 $E(\xi)\geq E(\eta)$.

常数作为只有一个可能值的离散型随机变量，其期望 $E(c) = c \cdot 1 = c$. 性质 2° 至 5° 都是积分中熟知的性质，不再一一证明.

如下定理给出数学期望一个更深刻的性质.

定理 1.5 若随机变量 ξ 有分布函数 $F(x)$, 则对于任何波雷尔函数 f 有

$$6^\circ \quad E\left[f(\xi)\right] = \int_{-\infty}^{+\infty} f(x)\,dF(x).$$

上式在如下意义下成立，如一方存在则另一方也存在，并且二者相等.

性质 6° 就是积分变换定理的 (1.14) 式，其中 (1.14) 式的左端关于 ξ 的概率分布 $\mathbf{F}$ 的积分就是关于其分布函数 $F(x)$ 的勒贝格 - 斯蒂尔吉斯积分. 但是却难于写出此定理的初等证明，现仅就如下特殊情形做一说明.

进一步假设随机变量 ξ 有密度函数 $p(x)$, 且 $f(x)$ 严格上升，有连续可微的反函数 $h(y)$. 此时，按 2.7.2 节定理 7.2 知 $\eta = f(\xi)$ 有密度 $p[h(y)]h'(y)$. 于是由定义中 (1.2) 式算得 η 的期望为

$$E(\eta) = \int_{-\infty}^{+\infty} yp[h(y)]h'(y)\,dy = \int_{-\infty}^{+\infty} f(x)p(x)\,dx.$$

最后一步用到变数替换 $x = h(y)$. 注意上述推导已对函数 f 添加了很强的限制. 不过，定理 1.5 的离散型情形的证明却不难，留作练习.

性质 6° 是计算随机变量函数之数学期望的有力工具，它在多元情形依然成立. 即当 $(\xi_1, \cdots, \xi_n)$ 有联合分布函数 $F(x_1, \cdots, x_n)$ 时，对任何 n 元波雷尔函数 f, 有

$$E[f(\xi_1, \cdots, \xi_n)] = \int \cdots \int_{\mathbb{R}^n} f(x_1, \cdots, x_n)\,dF(x_1, \cdots, x_n).$$

就是说，可直接用 $(\xi_1, \cdots, \xi_n)$ 的联合分布计算 $\eta = f(\xi_1, \cdots, \xi_n)$ 的数学期望，而不必先求 η 的概率分布.

有了性质 6°, 可分别对离散型与连续型导出定理 1.2. 例如当 ξ 与 η 分别有密度函数 $p_1(x)$ 与 $p_2(y)$ 时, 运用期望的性质 6° 可得

$$E(\xi\eta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp_1(x)p_2(y)\,dxdy$$

$$= \int_{-\infty}^{+\infty} xp_1(x)\,dx \int_{-\infty}^{+\infty} yp_2(y)\,dy = E(\xi)E(\eta).$$

离散型情形留作练习.

3.1.4 期望的极限性质 *

在概率论的理论及应用中, 经常会遇到随机变量序列 $\{\xi_n\}$ 之期望的极限问题, 即要考虑等式

$$\lim_n E(\xi_n) = E(\lim_n \xi_n) \tag{1.17}$$

成立的条件. 由于数学期望是积分, 故问题的实质是研究在什么条件下极限运算与积分运算可以交换次序. 这个问题在学习参变积分与函数项级数时就有过接触, 而实变函数论则针对实函数的积分系统地研究这个问题. 至于对抽象空间中可测函数的积分的讨论, 则是测度论中的内容. 本节将用尽可能简短通俗的语言叙述几个重要结论, 为不熟悉这一内容的读者扫清今后学习的障碍.

仍然以概率空间上随机变量对概率的积分 (即期望) 为例, 所得定理适用于一般可测函数对一般测度 (包括直线上的勒贝格测度) 的积分.

回忆 3.1.2 节关于概率测度积分的定义, 当 $\{\xi_n\}$ 是非负简单随机变量的上升序列时, 我们将其极限 $\xi = \lim_n \xi_n$ 的积分定义为

$$\int_\Omega \xi(\omega)p(d\omega) = \lim_n \int_\Omega \xi_n(\omega)p(d\omega),$$

这本身就肯定了此时极限与积分可以交换次序, 即 (1.17) 式成立. 下面先推广到一般非负随机变量列的情形.

定理 1.6 单调收敛定理 设随机变量序列 $\{\xi_n\}$ 满足条件

$$0 \le \xi_1(\omega) \le \xi_2(\omega) \le \cdots \le \xi_n(\omega) \uparrow \xi(\omega), \qquad \omega \in \Omega.$$

则 (1.17) 式成立.

证明: 对每个 $n \ge 1$, 设 $\{\eta_k^{(n)}\}$ 为定义 $E(\xi_n)$ 的简单随机变量列, 它满足

$$\begin{cases} 0 \le \eta_1^{(n)} \le \eta_2^{(n)} \le \cdots \le \eta_k^{(n)} \uparrow \xi_n, \\ \lim\limits_k E[\eta_k^{(n)}] = E(\xi_n). \end{cases} \tag{1.18}$$

令

$$\zeta_k = \bigvee_{i=1}^{k} \eta_k^{(i)}, \qquad k \ge 1, \tag{1.19}$$

则作为 k 个非负简单随机变量的最大值, 每个 ζ_k 均为非负简单随机变量. 再由 $\{\eta_k^{(i)}\}$ 对 k 非降知 $\{\zeta_k\}$ 也非降, 并且有

$$\begin{cases} \eta_k^{(n)} \le \zeta_k \le \xi_k, \\ E[\eta_k^{(n)}] \le E(\zeta_k) \le E(\xi_k), \end{cases} \qquad k \ge 1, \tag{1.20}$$

上式先固定 n, 令 $k \to \infty$, 用 (1.18) 可得

$$\begin{cases} \xi_n \le \lim\limits_k \zeta_k \le \xi \\ E(\xi_n) \le \lim\limits_k E(\zeta_k) \le \lim\limits_k E(\xi_k), \end{cases} \tag{1.21}$$

再令 $n \to \infty$ 就得到

$$\begin{cases} \lim\limits_k \zeta_k = \xi, \\ \lim\limits_k E(\zeta_k) = \lim\limits_n E(\xi_n). \end{cases} \tag{1.22}$$

至此我们得证 $\{\zeta_n\}$ 是非负简单随机变量的非降序列, 并且有 $\lim\limits_k \zeta_k = \xi$. 故用 ξ 之积分定义可得

$$E(\xi) = \lim_k E(\zeta_k),$$

联合 (1.22) 中第二个等式便得 (1.17) 式成立．证毕．

由于此定理只涉及非负情形，我们未对可积性提出要求．事实上，当涉及的积分值为正无穷时 (1.17) 式仍真．

定理 1.7 法都引理 设 $\{\xi_n\}$ 是一随机变量序列．

(1) 如果存在可积随机变量 σ, 使 $\xi_n \geq \sigma$, 则有

$$E(\varliminf_{n} \xi_n) \leq \varliminf_{n} E(\xi_n). \tag{1.23}$$

(2) 如果存在可积随机变量 τ, 使 $\xi_n \leq \tau$, 则有

$$E(\varlimsup_{n} \xi_n) \geq \varlimsup_{n} E(\xi_n). \tag{1.24}$$

证明: (1) 如果 $\varliminf_{n} E(\xi_n) = +\infty$, 则 (1.23) 式已真．以下设 $\varliminf_{n} E(\xi_n)$ 有限．取

$$\eta_n = \bigwedge_{k=n}^{\infty} (\xi_k - \sigma), \qquad n \geq 1.$$

则 $\{\eta_n\}$ 为非负随机变量的上升序列，它以 $\varliminf_{n}(\xi_n - \sigma)$ 为极限，故用定理 1.6 得

$$E\left[\varliminf_{n}(\xi_n - \sigma)\right] = \lim_{n} E\left[\bigwedge_{k=n}^{\infty} (\xi_k - \sigma)\right] \leq \varliminf_{n} E(\xi_n - \sigma). \tag{1.25}$$

消去有限的 $E(\sigma)$ 就推出 (1.23) 式．

(2) 对 $\{-\xi_n\}$ 用 (1) 便可得证 (1.24) 式，定理证毕．

我们用上述法都引理导出如下最具广泛应用的结果．

定理 1.8 勒贝格控制收敛定理 设 $\{\xi_n\}$ 为随机变量序列．如果存在可积随机变量 η 使 $|\xi_n| \leq \eta$, 且 $\lim_{n} \xi_n = \xi$, 则随机变量 ξ 可积，且有

$$\lim_{n} E(\xi_n) = E(\xi). \tag{1.26}$$

证明: $|\xi_n| \le \eta$ 蕴含 $|\xi| \le \eta$, 故 ξ 可积. 取 $\sigma = -\eta$ 与 $\tau = \eta$, 用法都引理便得到

$$E(\xi) \le \underline{\lim_n} E(\xi_n) \le \overline{\lim_n} E(\xi_n) \le E(\xi).$$

这说明极限 $\lim_n E(\xi_n)$ 存在且等于 $E(\xi)$, 定理至此得证.

有关极限与期望运算交换次序的这些定理, 在后面的有关特征函数与极限定理等内容中将会多次用到, 同时也是今后学习随机过程论的必备知识.

3.1.5 常见分布的期望

本节计算常见分布的数学期望. 注意我们有时从定义 (1.1) 或 (1.2) 式出发, 有时则是利用性质 3° 或 6° 来计算期望.

贝努里分布 随机变量 ξ 有分布: $P\{\xi = 0\} = q$, $P\{\xi = 1\} = p$. 则由 (1.1) 式立得 $E(\xi) = 0 \cdot q + 1 \cdot p = p$.

二项分布 设 ξ 服从二项分布, 则 ξ 可看成是 n 重贝努里试验中成功的次数. 令

$$\xi_k = \begin{cases} 1, & \text{若第 } k \text{ 次试验成功}, \\ 0, & \text{若第 } k \text{ 次试验失败}. \end{cases}$$

则 $\xi = \sum_{k=1}^{n} \xi_k$, 且对一切 k, ξ_k 都服从参数为 p 的贝努里分布, $E(\xi_k) = p$. 用性质 3° 便得 $E(\xi) = \sum_{k=1}^{n} E(\xi_k) = np$.

普阿松分布 可由 (1.1) 式直接算出普阿松分布随机变量的期望

$$E(\xi) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.$$

普阿松分布的唯一参数 λ 恰好是它的数学期望, 可见这种分布由其期望完全确定.

几何分布 从 (1.1) 式立得几何分布随机变量的期望

$$E(\xi) = \sum_{k=1}^{\infty} k q^{k-1} p = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} q^{k-1} p = \frac{1}{p}.$$

可见几何分布也由其数学期望完全确定.

均匀分布 设 ξ 服从区间 (a,b) 上的均匀分布, 由定义 (1.2) 式直接算得

$$E(\xi)=\int_{-\infty}^{+\infty}xp(x)\,dx=\int_a^b\frac{x}{b-a}\,dx=\frac{a+b}{2}.$$

注意 (1.2) 式恰为以 $p(x)$ 为线密度的不均匀棒的重心坐标. 对于 (a,b) 上的均匀棒, 其重心自应是区间的中点 $(a+b)/2$.

正态分布 先设随机变量 ξ 有 $N(0,1)$ 分布, 用 (1.2) 式应有

$$E(\xi)=\int_{-\infty}^{+\infty}\frac{x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx.$$

注意此积分绝对收敛, 而且被积函数为奇函数, 它在每个对称区间 $(-a,\ a)$ 上积分值为 0, 故 $E(\xi)=0$. 为求一般正态分布的期望, 令 $\eta=\sigma\xi+a$, 则 η 服从 $N(a,\sigma^2)$ 分布. 用期望的性质 $3^\circ, 2^\circ, 1^\circ$ 可得

$$E(\eta)=\sigma E(\xi)+a=\sigma\cdot 0+a=a.$$

就是说, $N(a,\sigma^2)$ 分布中第一个参数 a 恰为此分布的数学期望.

χ^2 分布 可以按 2.7.2 节 (7.5) 式给出的密度函数, 直接用定义中 (1.2) 式计算期望. 我们先对 $n=1$ 情形, 用性质 6° 计算 $E(\xi^2)$, 其中 ξ 是 $N(0,1)$ 分布随机变量, 便得自由度 $n=1$ 之 χ^2 分布的期望

$$\begin{aligned}E(\xi^2)&=\int_{-\infty}^{+\infty}\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}x^2e^{-\frac{x^2}{2}}dx\\&=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}\sqrt{2y}e^{-y}dy=\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right)=1.\end{aligned}$$

一般地, n 个自由度 χ^2 分布的期望 $E(\eta)=n$, 即期望值恰为其自由度, 这是决定此分布的唯一参数.

柯西分布 设随机变量 ξ 有密度函数

$$p(x)=\frac{1}{\pi}\frac{1}{1+x^2},$$

注意到

$$\int_{-\infty}^{+\infty} |x|p(x)\,dx = \int_{-\infty}^{+\infty} \frac{1}{\pi}\frac{|x|}{1+x^2}\,dx = +\infty,$$

故柯西分布的期望不存在. 在本书介绍的常见分布中, 数学期望不存在的仅柯西分布一种, 前面计算的各例均不存在这个问题, 未一一提及.

3.1.6 习题

1. 某人有 n 把外形相似的钥匙, 其中只有一把能打开门. 现一一试开, 直至打开门为止. 试对每次试毕 (1) 不放回; (2) 放回二情形求试开次数 ξ 的数学期望.

2. 设随机变量 ξ 有拉普拉斯分布, 其密度函数为

$$p(x) = \frac{1}{2\lambda}e^{-\frac{1}{\lambda}|x-\mu|}, \qquad (\lambda > 0),$$

求 $E(\xi)$.

3. 试求伽马分布 $\Gamma(\lambda, r)$ 的数学期望.

4. 在长为 a 的线段上任意独立地取 n 个点, 求相距最远的两点间距离的期望.

5. 设 ξ 为非负整值随机变量, 其数学期望存在, 求证

$$E(\xi) = \sum_{i=1}^{\infty} P\{\xi \geq i\}.$$

6. 设 $F(x)$ 为某非负随机变量的分布函数, 试证对任 $s > 0$ 有

$$\int_0^{+\infty} x^s\,dF(x) = \int_0^{+\infty} sx^{s-1}[1-F(x)]\,dx.$$

7. 设随机变量 ξ 的分布函数为 $F(x)$, 且 ξ 的期望存在. 求证

$$E(\xi) = \int_0^{+\infty} [1-F(x)]\,dx - \int_{-\infty}^{0} F(x)\,dx.$$

8. 设 ξ 服从 $N(a, \sigma^2)$ 分布, 求 $\eta = e^{\frac{a^2-2a\xi}{2\sigma^2}}$ 的数学期望.

9. 设 ξ 为柯西分布 $C(1,0)$ 随机变量，计算 $E(|\xi| \wedge 1)$.

10. 设随机变量 ξ,η 独立同 $N(a,\sigma^2)$ 分布，求证

$$E(\xi \vee \eta) = a + \frac{\sigma}{\sqrt{\pi}}.$$

11. 设随机变量 ξ,η 相互独立，同服从 $g(k;p)$ 分布，试求 $E(\xi\vee\eta)$.

12. 袋中有 r 个红球与 b 个黑球，现从中任取 n 个球，求其中红球个数的数学期望.

13. 参加集会的 n 个人将他们的帽子混放在一起，会后每人任取一顶帽子戴上，以 ξ 表示他们中戴对自己帽子的人的个数，求 $E(\xi)$.

14. 设袋中有 2^n 个球，其中编号为 k 的球各 $\binom{n}{k}$ 个 $(k=0,1,\cdots,n)$. 现不放回地从袋中任取 m 个 $(m<2^n)$，求这些球上编号之和的数学期望.

15. 袋中有 r 个红球与 b 个黑球，现任意一一取出，直至取到红球为止，求取球次数的数学期望.

3.2 其它数字特征

3.2.1 两个引理

以下介绍描述概率分布其它特性的数字特征，其中包括反映随机变量取值分散程度的数字特征，描述随机变量间相依程度的数字特征等. 这些数字特征实际上都是随机变量函数的数学期望. 我们的主要目标是，选择适当的 f，使 $E[f(\xi)]$ 能刻画 ξ 之概率分布的某种特性. 由于我们选用的 f 常常是二次函数，故有必要先介绍平方可积随机变量族及有关的性质.

设 $(\Omega,\mathcal{F},P)$ 为给定的概率空间，其上平方可积随机变量 ξ 的全体记为

$$L^2 = L^2(\Omega,\mathcal{F},P) = \{\xi : E(\xi^2) < +\infty\}. \tag{2.1}$$

利用初等不等式 $(\xi+\eta)^2 \le 2(\xi^2+\eta^2)$ 及期望的性质知，若 ξ，$\eta \in L^2$ 则它们的任意线性组合 $a\xi + b\eta \in L^2$.

由于在大多数情况下只是研究事件的概率或随机变量的积分(期望),此时所涉及的命题都允许有零概率的例外.例如对连续型场合,我们不区分"非负值随机变量"与"正值随机变量",或者说,在不计零概率差异意义下可以认为二者相同.称这种可以有零概率例外集的命题为**几乎必然成立**的,简记为 $a.s.$ 成立.现在已准备好介绍空间 L^2 的两条有用的性质.

引理 2.1 $E(\xi^2)=0$ 的充分必要条件是 $\xi=0.\ a.s.$

证明: 充分性不待证.为证必要性,我们用反证法.设 $P\{|\xi|>0\}>0$,则存在自然数 k 使

$$P\{|\xi|>\frac{1}{k}\}=\varepsilon>0.$$

以 $F(x)$ 表 ξ 的分布函数,则有

$$E(\xi^2)=\int_{\Re}x^2\,dF(x)\geq\int_{\{|x|>\frac{1}{k}\}}x^2\,dF(x)$$

$$>\frac{1}{k^2}P\left\{|\xi|>\frac{1}{k}\right\}=\frac{\varepsilon}{k^2}>0.$$

这与 $E(\xi^2)=0$ 矛盾,引理得证.

引理 2.2 柯西-施瓦兹不等式 设 $\xi,\eta\in L^2$,则

$$[E(\xi\eta)]^2\leq E(\xi^2)E(\eta^2). \tag{2.2}$$

上式等号成立,当且仅当存在实数 t_0 使

$$\eta=t_0\xi.\qquad a.s. \tag{2.3}$$

证明: 由 L^2 对线性运算封闭知 $\eta-t\xi\in L^2$,于是

$$u(t)=E(\eta-t\xi)^2=t^2E(\xi^2)-2tE(\xi\eta)+E(\eta^2)$$

是 t 的实值二次函数,且对一切 t 有 $u(t)\geq 0$,故其判别式 $\Delta=[E(\xi\eta)]^2-E(\xi^2)E(\eta^2)\geq 0$,此即得证 (2.2) 式.等号成立当且仅当 $\Delta=0$,当且仅当有唯一的 t_0 使 $u(t_0)=E(\eta-t_0\xi)^2=0$,再用引理 2.1 知,当且仅当存在 t_0 使 (2.3) 式成立.证毕.

3.2.2 方差

本节给出描述随机变量 ξ 取值相对于均值 $E(\xi)$ 的分散程度的数字特征. 以离散型为例, ξ 的每个可能值 x_i 与其均值 $E(\xi)$ 有偏差 $x_i - E(\xi)$, 如果直接把这些偏差对 ξ 的一切可能值累加起来, 那么正负相抵消将造成总和不能反映 ξ 取值的分散程度. 于是改用各偏差的平方和来代表总偏差. 应用中为保持单位的一致, 有时再取这个总偏差的算术根.

我们只考虑空间 L^2 中的随机变量 ξ, 以保证所涉及的积分 $E|\xi|^2$ 与 $E|\xi|$ 有限. 事实上, 当 $\xi \in L^2$ 时, 利用初等不等式 $|\xi| \le 1 + |\xi|^2$ 立得 ξ 的期望有限.

定义 2.1 对于 $\xi \in L^2$, 称

$$D(\xi) = E[\xi - E(\xi)]^2 \tag{2.4}$$

为 ξ (及其概率分布) 的**方差**, 而称 $\sqrt{D(\xi)}$ 为其**标准差**.

方差有如下基本性质

定理 2.3 设 $\xi \in L^2$, 则有

1° $D(\xi) = 0$ 当且仅当 $\xi = E(\xi)$, $a.s.$;

2° $D(\xi) = E(\xi^2) - [E(\xi)]^2$;

3° $D(c\xi) = c^2 D(\xi)$;

4° $f(c) = E(\xi - c)^2$ 当且仅当 $c = E(\xi)$ 时取到最小值.

证明: 运用定义中 (2.4) 式及引理 2.1 立得性质 1°. 将 (2.4) 式右方展开并用期望的线性性得证 2°. 由 2° 直接计算便得 3°. 最后由

$$f(c) = E[\xi - E(\xi) + E(\xi) - c]^2$$

$$= E[\xi - E(\xi)]^2 + [E(\xi) - c]^2$$

可见 $f(c) \geq D(\xi)$, 当且仅当 $c = E(\xi)$ 时取等号. 定理证毕.

(2.4) 式在离散型与连续型场合的表达式分别为

$$D(\xi) = \sum_i [x_i - E(\xi)]^2 p_i, \tag{2.5}$$

$$D(\xi) = \int_{-\infty}^{+\infty} (x - E(\xi))^2 p(x)\, dx. \tag{2.6}$$

注意, 若一不均匀棒以 $p(x)$ 为其线密度, 则 $D(\xi)$ 是此棒以过质心的直线为轴的转动惯量. 方差的性质 4° 表明, 如果转动轴放在其它直线 $x = c$ 处, 则转动惯量 $f(c)$ 必然变大.

可以通过 (2.5) 或 (2.6) 式计算常见分布的方差, 也可以运用方差的性质 2° 来计算. 以下只算三个例子.

例 2.1 普阿松分布的方差 设随机变量 ξ 服从 $p(k;\lambda)$ 分布, 已有 $E(\xi) = \lambda$. 下面用性质 2° 求 $D(\xi)$, 只要再算 $E(\xi^2)$. 由期望的性质 6° 可得

$$E(\xi^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} [k(k-1) + k] \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \lambda^2 \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} + \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda^2 + \lambda.$$

至此得普阿松分布的方差 $D(\xi) = E(\xi^2) - [E(\xi)]^2 = \lambda$.

例 2.2 正态分布的方差 设 ξ 服从 $N(a, \sigma^2)$ 分布, 已有 $E(\xi) = a$, 用 (2.6) 式得

$$D(\xi) = \int_{-\infty}^{+\infty} (x-a)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx,$$

作变数替换 $u = (x-a)/\sigma$, 正好化为 3.1.5 节求 χ^2 分布期望时算过的积分. 于是得到

$$D(\xi) = \sigma^2 \int_{-\infty}^{+\infty} \frac{u^2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} = \sigma^2.$$

可见正态分布 $N(a,\sigma^2)$ 的方差是此分布的第二个参数 σ^2, 从而正态分布由它的期望与方差完全确定. 回忆第二章介绍正态分布时, 曾指出 σ^2 是此分布的形状参数, σ^2 越大正态密度曲线越平缓, ξ 的取值越分散. 现又算得 $D(\xi)=\sigma^2$, 可见方差确实反映了 ξ 取值的分散程度.

例 2.3 $\chi^2(n)$分布的方差 已经有此分布的期望 $E(\xi)=n$, 再由期望的性质 6° 得

$$\begin{aligned}
E(\xi^2) &= \int_{-\infty}^{+\infty} x^2 p(x)\,dx \\
&= \int_0^{+\infty} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}+1} e^{-\frac{x}{2}}\,dx \\
&= \frac{4}{\Gamma(\frac{n}{2})}\int_0^{+\infty} y^{\frac{n}{2}+1}e^{-y}\,dy = \frac{4}{\Gamma(\frac{n}{2})}\Gamma\left(\frac{n}{2}+2\right) \\
&= 4\left(\frac{n}{2}+1\right)\cdot\frac{n}{2} = n^2+2n.
\end{aligned}$$

再利用方差性质 2° 便得 $D(\xi)=n^2+2n-n^2=2n$. $\chi^2(n)$ 分布仅含一个参数, 即自由度 n, 它既是此分布的期望, 又是此分布方差的一半.

注意在以上三个求方差的例子中, 由于只涉及非负函数的求和或积分, 我们采用直接计算 $E(\xi^2)$ 或 $D(\xi)$ 的方法, 算得结果是有限的, 则同时证明了 $\xi\in L^2$. 只有柯西分布, 由于其期望不存在, 方差自然也不存在. 其余常见分布的方差可参见附表 I(常见分布表).

由期望的序性知 $D(\xi)\geq 0$, 而方差的性质 1° 又表明条件 $D(\xi)>0$ 仅排除了 ξ 恒等于常数这个退化情形. 以下称使 $D(\xi)>0$ 的随机变量为**非退化随机变量**.

定义 2.2 设 ξ 为 L^2 中非退化随机变量, 称

$$\xi^*=\frac{\xi-E(\xi)}{\sqrt{D(\xi)}} \tag{2.7}$$

为 ξ 的**标准化随机变量**.

易见 $E(\xi^*)=0$ 且 $D(\xi^*)=1$.

例如，正态分布 $N(a,\sigma^2)$ 随机变量的标准化为 $(\xi-a)/\sigma$, 参数为 λ 的普阿松随机变量的标准化为 $(\xi-\lambda)/\sqrt{\lambda}$ 等等.

3.2.3 协方差阵

以下考虑多维概率分布的数字特征. 首先, n 维随机变量 $(\xi_1,\cdots,\xi_n)$ 的每个分量 ξ_i 都有自己的数字特征. 正如在 3.1.3 节开头所指出的, 既可以用边缘分布，也可以用联合分布求出分量的数字特征. 比如二维正态分布，由于已得到两个边缘分布分别为 $N(a_1,\sigma_1^2)$ 与 $N(a_2,\sigma_2^2)$, 故参数 a_1 与 a_2 分别是两个分量的数学期望，而 σ_1^2 与 σ_2^2 则是它们各自的方差. 下面着手研究描写各分量相依关系的数字特征.

定义 2.3 设随机变量 $(\xi_1,\cdots,\xi_n)$ 的每个分量 $\xi_i\in L^2$, 称

$$\operatorname{cov}(\xi_i,\xi_j)=E[(\xi_i-E(\xi_i))(\xi_j-E(\xi_j))] \tag{2.8}$$

为 ξ_i 与 ξ_j 的**协方差**, 而将 $n\times n$ 方阵 $B=[b_{ij}]$, 其中 $b_{ij}=\operatorname{cov}(\xi_i,\xi_j)$, 称为这个随机变量的**协方差阵**.

注意 (2.8) 式的右方不过是随机向量 $(\xi_1,\cdots,\xi_n)$ 的某个函数的期望，它应当按 $(\xi_1,\cdots,\xi_n)$ 的联合分布用期望的性质 6° 求出. 再者，协方差阵的主对角线元 $b_{ii}=D(\xi_i)$. 由于假定 $\xi_i\in L^2$, 由柯西 - 施瓦兹不等式知每个 b_{ij} 有限. 今后我们将逐步说明，(2.8) 式定义的协方差确实描述了 ξ_i 与 ξ_j 的某种相依程度.

定理 2.4 协方差阵 B 有如下性质

1° $\operatorname{cov}(\xi_i,\xi_j)=E(\xi_i\xi_j)-E(\xi_i)E(\xi_j)$;

2° B 为对称半正定阵;

3° $D(\xi_i+\xi_j)=D(\xi_i)+D(\xi_j)+2\operatorname{cov}(\xi_i,\xi_j)$.

证明: 将协方差定义 (2.8) 式右方展开便得证 1°. 其次，对一切 $i,j=1,\cdots,n$ 显然有对称性 $b_{ij}=b_{ji}$ 成立. 再者，对任意 n 个实数

$a_1, \cdots, a_n$, 利用数学期望的线性与序性便得

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j b_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\left[(\xi_i - E(\xi_i))(\xi_j - E(\xi_j))\right]$$

$$= E\left[\sum_{i=1}^{n} a_i(\xi_i - E(\xi_i))\right]^2 \geq 0,$$

这表明 B 是半正定阵，性质 2° 得证. 最后，直接按定义 (2.4) 式算 $\xi_i + \xi_j$ 的方差，并用期望的线性可得

$$D(\xi_i + \xi_j) = E[\xi_i + \xi_j - E(\xi_i) - E(\xi_j)]^2$$

$$= E[\xi_i - E(\xi_i)]^2 + E[\xi_j - E(\xi_j)]^2 + 2E[(\xi_i - E(\xi_i))(\xi_j - E(\xi_j))]$$

$$= D(\xi_i) + D(\xi_j) + 2\mathrm{cov}(\xi_i, \xi_j).$$

此即性质 3°. 定理得证.

例 2.4 回到 2.5.3 节的例 5.5, 试对 (1) 有放回； (2) 无放回两种情形，求两次取数结果 ξ 与 η 的协方差.

解: 由于边缘分布均为贝努里分布，故有

$$E(\xi) = E(\eta) = \frac{2}{5}, \quad D(\xi) = D(\eta) = \frac{6}{25},$$

下面计算 $\mathrm{cov}(\xi, \eta)$, 在情形 (1), 我们有

$$E(\xi\eta) = 0 \cdot 0 \cdot \frac{9}{25} + 0 \cdot 1 \cdot \frac{6}{25} + 1 \cdot 0 \cdot \frac{6}{25} + 1 \cdot 1 \cdot \frac{4}{25} = \frac{4}{25}.$$

故用协方差性质 1°

$$\mathrm{cov}(\xi, \eta) = E(\xi\eta) - E(\xi)E(\eta) = \frac{4}{25} - \frac{4}{25} = 0.$$

在情形 (2), 由于

$$E(\xi\eta) = 0 \cdot 0 \cdot \frac{6}{20} + 0 \cdot 1 \cdot \frac{6}{20} + 1 \cdot 0 \cdot \frac{6}{20} + 1 \cdot 1 \cdot \frac{2}{20} = \frac{2}{20},$$

所以有

$$\mathrm{cov}(\xi, \eta) = E(\xi\eta) - E(\xi)E(\eta) = \frac{2}{20} - \frac{4}{25} = -\frac{3}{50}.$$

例 2.5　二维正态分布的协方差阵　设 (ξ,η) 服从二维正态分布. 已有 $E(\xi)=a_1$, $E(\eta)=a_2$, $b_{11}=D(\xi)=\sigma_1^2$, $b_{22}=D(\eta)=\sigma_2^2$. 下算协方差 $b_{12}=\mathrm{cov}(\xi,\eta)$. 直接用定义 (2.8) 式的连续型表达式并作变数替换

$$u=\frac{x-a_1}{\sigma_1},\quad v=\frac{y-a_2}{\sigma_2}$$

可得

$$\mathrm{cov}(\xi,\eta)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{\sigma_1\sigma_2uv}{2\pi\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}(u^2-2ruv+v^2)}dudv.$$

配方后再令

$$s=\frac{u-rv}{\sqrt{1-r^2}},\quad t=v$$

便推出

$$\mathrm{cov}(\xi,\eta)=\frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\sqrt{1-r^2}st+rt^2)e^{-\frac{s^2+t^2}{2}}dsdt$$

$$=\sigma_1\sigma_2\left(0+r\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds\int_{-\infty}^{+\infty}\frac{t^2}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\right)=r\sigma_1\sigma_2.$$

至此我们算得二维正态分布的协方差阵为

$$B=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}.$$

本节开头曾指出, 我们的目的是引进描述随机变量间相依程度的数字特征. 那么, 由 (2.8) 式定义的协方差是否符合这一要求呢? 回顾上述两个例题的结果. 在例 2.4 的有放回情形 (1) 中, 两次取出的数 ξ 与 η 是相互独立的, 此时恰有 $\mathrm{cov}(\xi,\eta)=0$. 而在不放回情形 (2), 则 $\mathrm{cov}(\xi,\eta)\neq 0$. 在二维正态分布中, 我们算得协方差 $\mathrm{cov}(\xi,\eta)=r\sigma_1\sigma_2$, 但前已指出, 当参数 $r=0$ 时 ξ 与 η 独立, 此时协方差也是 0; 反之, 当 $\mathrm{cov}(\xi,\eta)\neq 0$ 时 $r\neq 0$, ξ 与 η 不独立. 两个例子有共同的特点: 当两个随机变量独立时 (这应当是最松散的相依关系), 协方差为 0. 这至少是协方差反映随机变量间相依程度的一个例证.

3.2.4 相关系数

协方差是一个有单位的量，比如 (ξ,η) 分别表示人群中任意一人的身高与体重. 如果 ξ 以米为单位，η 以公斤为单位，则算出的协方差以"米 · 公斤"为单位. 若改用其它单位，则协方差的数值可能有很大变化. 因此，在有些场合需要用不依赖度量单位取法的量来表示这一数字特征.

定义 2.4 设 ξ,η 为 L^2 中非退化随机变量，称

$$r = \frac{\mathrm{cov}(\xi,\eta)}{\sqrt{D(\xi)}\sqrt{D(\eta)}} \tag{2.9}$$

为 ξ 与 η 的**相关系数**. 当 $r=0$ 时，称 ξ 与 η**不相关**.

由定义立得

$$r = E\left[\frac{\xi - E(\xi)}{\sqrt{D(\xi)}} \cdot \frac{\eta - E(\eta)}{\sqrt{D(\eta)}}\right] = E(\xi^*\eta^*). \tag{2.10}$$

就是说，相关系数是标准化后的协方差.

在上述例 2.4 的情形 (2) 中，两次取出的数字之相关系数为

$$r = \frac{-\frac{3}{50}}{\sqrt{\frac{6}{25}}\sqrt{\frac{6}{25}}} = -\frac{1}{4}.$$

而在例 2.5 中，二维正态分布两分量的相关系数恰好是参数 r. 以下三个定理给出相关系数的重要性质.

定理 2.5 对空间 L^2 中的非退化随机变量 ξ,η 而言，其相关系数 $r\in[-1,1]$； $r=1$ 当且仅当 $\eta^*=\xi^*, a.s$； $r=-1$ 当且仅当 $\eta^*=-\xi^*$, $a.s.$

证明: 注意到 (2.10) 式, 对 ξ^*,η^* 用柯西 - 施瓦兹不等式 (2.2) 可得

$$|r| = |E(\xi^*\eta^*)| \le \sqrt{D(\xi^*)D(\eta^*)} = 1.$$

而且 $|r| = 1$ 当且仅当存在 t_0 使 $\eta^* = t_0\xi^*, a.s.$ 至此只要再证明 $t_0 = r$ 即完成了定理的证明. 事实上, 此时有

$$r = E(\xi^*\eta^*) = E(t_0\xi^{*2}) = t_0E(\xi^{*2}) = t_0.$$

定理得证.

此定理表明, 当 $|r| = 1$ 时, 存在常数 a,b 使 $\eta = a + b\xi, a.s.$ 且当 $r = 1$ 时, $b > 0$, ξ 与 η 为完全正线性相关; 当 $r = -1$ 时, $b < 0$, ξ 与 η 完全负线性相关.

定理 2.6 对于空间 L^2 中非退化随机变量而言, 如下四命题等价:

(1) ξ,η 不相关; (2) $\mathrm{cov}(\xi,\eta) = 0$;
(3) $E(\xi\eta) = E(\xi)E(\eta)$; (4) $D(\xi+\eta) = D(\xi) + D(\eta)$.

证明: 由定义 2.4 及协方差性质 1° 及 3° (见定理 2.2), 立即得证这四个命题的等价性.

定理 2.7 对于空间 L^2 中非退化随机变量而言, 若 ξ 与 η 独立, 则 ξ 与 η 不相关.

证明: 在 3.1.2 节定理 1.2 中已证明, 当 ξ,η 独立时, 上述不相关的等价条件 (3) 成立. 证毕.

对于二元正态随机变量, 此定理的逆也真. 即 $r = 0$ 时, 二分量 ξ 与 η 独立. 但对一般的 ξ 与 η, 定理 2.7 的逆不真, 即不相关随机变量不一定相互独立. 且看下面的例.

例 2.6 设 θ 为区间 $(0,2\pi)$ 上均匀分布的随机变量, $\xi = \cos\theta, \eta = \cos(\theta + a)$, 其中 a 为常数. 不难算出

$$E(\xi) = E(\eta) = 0, \quad D(\xi) = D(\eta) = \frac{1}{2},$$

$$E(\xi\eta)=\int_0^{2\pi}\frac{1}{2\pi}\cos x\cos(x+a)\,dx=\frac{1}{2}\cos a.$$

故 ξ 与 η 的相关系数

$$r=\frac{E(\xi\eta)}{\sqrt{D(\xi)}\sqrt{D(\eta)}}=\cos a.$$

于是,

(1) 当 $a=0$ 时 $r=1$, 此时 $\eta=\cos\theta=\xi$.

(2) 当 $a=\pi$ 时 $r=-1$, 此时 $\eta=\cos(\theta-\pi)=-\cos\theta=-\xi$.

(3) 当 $a=-\pi/2$ 时 $r=0$ 此时 $\xi=\cos\theta,\eta=\sin\theta$. 显然 ξ 与 η 不相关, 但是它们却不独立. 我们严格证明此时 ξ,η 不满足如下独立定义:

$$P\{\xi\in B_1,\eta\in B_2\}=P\{\xi\in B_1\}P\{\eta\in B_2\},\ 任\ B_1,B_2\in\mathcal{B}.$$

事实上, 可取 $B_1=B_2=(0,1/2)$, 则有

$$P\{\eta\in B_2\}=P\left\{0<\sin\theta<\frac{1}{2}\right\}=P\left\{\theta\in\left(0,\frac{\pi}{6}\right)\cup\left(\frac{5}{6}\pi,\pi\right)\right\}$$

$$=\frac{1}{6}=P\left\{0<\cos\theta<\frac{1}{2}\right\}=P\{\xi\in B_1\}.$$

但因 $\{\xi\in B_1,\eta\in B_2\}=\varnothing$, 故

$$P\{\xi\in B_1,\eta\in B_2\}=0\neq P\{\xi\in B_1\}P\{\eta\in B_2\},$$

即 ξ 与 η 不是相互独立的.

结合对定理 2.5 的讨论可知, 相关系数 r 原来是反映 ξ 与 η 线性相关程度的数字特征. 当 $|r|$ 取到其最大值 1 时, ξ 与 η 以概率 1 成线性关系. 当 $|r|$ 取其最小值 0 时, ξ 与 η 不相关, 即不存在线性关系. 但此时可以有其它相依关系, 所以不一定是相互独立的. 例如对上例中的 $\xi=\cos\theta$ 与 $\eta=\sin\theta$, 它们不相关只是不存在线性关系, 但是有非线性关系 $\xi^2+\eta^2=1$, 即它们的值共存于单位圆周上.

3.2.5 条件数学期望与最优预测

在 2.6.1 节我们已介绍了随机变量的条件分布. 一般地, 对任何有正概率的事件 B, 定义随机变量 ξ 关于事件 B(即给定 B 已发生) 的条件分布函数为

$$F(x|B) = P\{\xi < x|B\} = \frac{P(\xi < x, B)}{P(B)}.$$

注意对固定的 B,$F(x|B)$ 也是一个分布函数, 因此可以考虑对它的勒贝格 - 斯蒂尔吉斯积分.

定义 2.5 如果相应的积分绝对收敛, 则称

$$E(\xi|B) = \int_{-\infty}^{+\infty} x\,dF(x|B) \tag{2.11}$$

为已知 B 发生后 ξ 的**条件数学期望**. 特别取 $B = \{\eta = y\}$, 称 $E\{\xi|\eta = y\}$ 为给定 $\eta = y$ 时 ξ 的条件期望.

对离散型与连续型的随机向量 (ξ,η), $E(\xi|\eta = y)$ 分别化为

$$E(\xi|\eta = y_j) = \sum_i x_i P\{\xi = x_i|\eta = y_j\}, \tag{2.12}$$

$$E(\xi|\eta = y) = \int_{-\infty}^{+\infty} x\,p(x|\eta = y)\,dx. \tag{2.13}$$

由于当 ξ,η 相互独立时, 条件分布等于无条件分布, 此时条件期望 $E(\xi|\eta = y)$ 与无条件期望 $E(\xi)$ 相等. 再者, 与无条件期望相似, 对随机变量的波雷尔函数 $g(\xi)$, 可用期望的性质 6° 算 $g(\xi)$ 关于 $\eta = y$ 的条件期望

$$E[g(\xi)|\eta = y] = \int_{-\infty}^{+\infty} g(x)\,dF(x|\eta = y).$$

特别，当 ξ 平方可积时，取 $g(\xi)=[\xi-E(\xi|\eta=y)]^2$，则得到给定 $\eta=y$ 后 ξ 的**条件方差**

$$D(\xi|\eta=y)=E\left\{(\xi-E[\xi|\eta=y])^2|\eta=y\right\}. \tag{2.14}$$

例如，对于单位圆内均匀分布的随机向量 (ξ,η)，当 $|y|\geq 1$ 时，由于 η 的边缘分布 $p_2(y)=0$，故在 $\eta=y$ 条件下 ξ 的条件分布无定义，当然也没有条件期望. 当 $|y|<1$ 时，给定 $\eta=y$ 后 ξ 服从区间 $(-\sqrt{1-y^2},\sqrt{1-y^2})$ 上的均匀分布 (参见 2.6.1 节)，故给定 $\eta=y\in(-1,1)$ 时 ξ 的条件期望与条件方差分别是

$$E(\xi|\eta=y)=0,\quad D(\xi|\eta=y)=\frac{1-y^2}{3}. \tag{2.15}$$

又如二维正态随机变量 (ξ,η)，由 2.6.1 节的 (6.4) 式知，给定 $\eta=y$ 时 ξ 的条件期望与条件方差分别是

$$E(\xi|\eta=y)=a_1+r\frac{\sigma_1}{\sigma_2}(y-a_2), \tag{2.16}$$

$$D(\xi|\eta=y)=\sigma_1^2(1-r^2).$$

一般地说，给定 $\eta=y$ 后 ξ 的条件期望应当是 η 的可能值 y 的函数，记之为 $\varphi(y)=E(\xi|\eta=y)$. 再将 y 用 η 代回，就得到一个随机变量 $\varphi(\eta)$，相应地记 $\varphi(\eta)=E(\xi|\eta)$. 就是说， $E(\xi|\eta)$ 是随机变量 η 的这样一个函数，当 $\eta=y$ 时， $E(\xi|\eta)$ 对应的值应是由 (2.12) 或 (2.13) 式定义的 $E(\xi|\eta=y)$. 我们称 $E(\xi|\eta)$ 为 ξ 关于 η 的条件期望. 同理，可定义 $E(\eta|\xi)$ 以及 $D(\xi|\eta)$, $D(\eta|\xi)$.

条件数学期望 $E(\xi|\eta)$ 在概率论的理论与应用上起着重要的作用. 它具备与初等数学期望 $E(\xi)$ 相类似的线性，序性等一序列性质. 限于篇幅，在此不能详述. 我们只介绍如下初等概率论中有用的性质.

定理 2.8 全期望公式 设 (ξ,η) 为联合分布的随机变量， $g(\xi)$ 为可积随机变量，则有

$$E\{E[g(\xi)|\eta]\}=E[g(\xi)]. \tag{2.17}$$

证明: 假定 (ξ,η) 是连续型的，有联合密度函数 $p(x,y)$，那么对 $\varphi(\eta)=E[g(\xi)|\eta]$ 用期望性质 6° 可得

$$
\begin{aligned}
E[E\{g(\xi)|\eta\}] &= \int_{-\infty}^{+\infty} E[g(\xi)|\eta=y]p_2(y)\,dy \\
&= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} g(x)p(x|\eta=y)\,dx\right]p_2(y)\,dy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)p(x,y)\,dxdy = E[g(\xi)].
\end{aligned}
$$

(2.17) 式得证．离散型情形证明类似，留作练习．定理证毕．

全期望公式 (2.17) 表明，条件期望 (作为一个随机变量) 的期望等于无条件期望．如果将这个等式反过来，它就象全概公式一样有着广泛的应用．概括地说，当我们需要求较复杂的随机变量 $g(\xi)$ 的期望时，可以借助于另一适当的随机变量 η，化无条件期望为条件期望，往往可以使问题容易解决．

例 2.7 巴格达窃贼问题 一窃贼被关在有 3 个门的地牢中．其中第 1 个门通向自由，出这个门后走 3 个小时便回到地面；第 2 个门通向一个地道，在此地道中走 5 个小时将返回地牢；第 3 个门通向一个更长的地道，沿这个地道走 7 个小时也回到地牢．如果窃贼每次选择 3 个门可能性总相等，求他为获自由而奔走的平均时间长．

解: 设窃贼需走 ξ 个小时到达地面，并设 η 代表窃贼每次对 3 个门的选择，η 各以 1/3 的可能性取值 1, 2, 3．运用全期望公式得

$$E(\xi)=E[E(\xi|\eta)]=\sum_{i=1}^{3}E(\xi|\eta=i)P\{\eta=i\}.$$

注意到

$$E(\xi|\eta=1)=3,\quad E(\xi|\eta=2)=5+E(\xi),$$

$$E(\xi|\eta=3)=7+E(\xi).$$

其中, 例如 $E(\xi|\eta=2)=5+E(\xi)$ 是因为, 当他选第 2 个门走出 5 个小时以后, 他将回到地牢, 处境与开始时完全一样. 连同 $P\{\eta=i\}=1/3$ 代回便可得到

$$E(\xi)=\frac{1}{3}[3+5+E(\xi)+7+E(\xi)],$$

于是有 $E(\xi)=15$ (小时). 即平均地说, 窃贼将在 15 小时后获得自由.

例 2.8 设随机变量 ξ 有密度函数

$$p_1(x)=\lambda^2 x\, e^{-\lambda x}, \qquad x>0.$$

随机变量 η 在区间 $(0,\xi)$ 上均匀分布. 试求 $E(\eta|\xi)$ 及 $E(\eta)$.

解: 因 $x>0$ 时, $E(\eta|\xi=x)=x/2$, 故 $E(\eta|\xi)=\xi/2$. 再用全期望公式 (2.17) 得

$$E(\eta)=E[E(\eta|\xi)]=E\left(\frac{\xi}{2}\right)=\frac{1}{2}\int_0^{+\infty}\lambda^2 x^2\, e^{-\lambda x}\,dx,$$

$$=\frac{1}{2\lambda}\int_0^{+\infty}u^2\, e^{-u}\,du=\frac{1}{2\lambda}\Gamma(3)=\frac{1}{\lambda}.$$

我们在习题 2.6.3 的第 5 题中求过 η 的密度函数, η 应服从以 λ 为参数的指数分布. 据此也可得相同的结果 $E(\eta)=1/\lambda$.

例 2.9 将成功率为 p 的贝努里试验独立地重复 η 次. η 是与每次试验结果相独立的随机变量, 它服从参数为 λ 的普阿松分布. 试求其中成功次数 ξ 的数学期望.

解: 对每个自然数 k, 令

$$\xi_k=\begin{cases}1, & \text{若第 } k \text{ 次试验成功},\\ 0, & \text{若第 } k \text{ 次试验失败}.\end{cases}$$

则 $\{\xi_k\}$ 为独立的随机变量序列, 它们同服从贝努里分布, 且有 $\xi=\sum_{k=1}^{\eta}\xi_k$. 问题化为求随机多个独立同分布随机变量之和的数学期望.

我们介绍以下一般结果:

定理 2.9 随机多个随机变量之和的期望 假设 $\{\xi_k\}$ 为相互独立的随机变量序列, 它们都与某个可积随机变量 ξ 同分布. 如果 η 为取值自然数的随机变量, $E(\eta)$ 有限且 η 与每个 ξ_k 独立, 那么我们有

$$E\left(\sum_{k=1}^{\eta}\xi_k\right)=E(\eta)E(\xi). \tag{2.18}$$

证明: 利用全期望公式 (2.17) 便得证

$$\begin{aligned}
E\left[\sum_{k=1}^{\eta}\xi_k\right] &= E\left[E\left(\sum_{k=1}^{\eta}\xi_k|\eta\right)\right]\\
&= \sum_{n=1}^{\infty}P(\eta=n)E\left(\sum_{k=1}^{n}\xi_k|\eta=n\right)\\
&= \sum_{n=1}^{\infty}\left[P(\eta=n)\sum_{k=1}^{n}E(\xi_k)\right]\\
&= \sum_{n=1}^{\infty}nP\{\eta=n\}E(\xi)=E(\eta)E(\xi).
\end{aligned}$$

运用 (2.18) 式立得例 2.9 的解为

$$E(\xi)=E\left(\sum_{k=1}^{n}\xi_k\right)=E(\eta)E(\xi_k)=\lambda p.$$

这个结果与 2.3.2 节的例 3.2 中导出的普阿松分布在随机选择下不变是一致的.

关于随机多个随机变量之和的方差, 在定理 2.9 的条件下有如下计算公式:

$$D\left(\sum_{k=1}^{\eta}\xi_k\right)=E(\eta)D(\xi)+D(\eta)[E(\xi)]^2. \tag{2.19}$$

我们将在下节运用母函数给出此式的一个证明.

条件数学期望可用来解决一类最优预测问题, 即联合分布随机向量之各分量间的预测问题. 比如人的身高 ξ 与体重 η, 在一些青年中

有一个流行的公式是 $\eta = \xi - 110$, 其中 ξ 以厘米为单位，而 η 以公斤为单位. 可以拿它作为用身高预测体重的公式. 但是，为了评价这个预测公式的好坏，就要提出一个标准，要明确追求什么样的目标. 没有目标函数的盲目预测显然是毫无意义的. 采用“均方误差达到最小”作为目标，我们给出如下结果.

定理 2.10 设 ξ 与 η 是联合分布着的平方可积随机变量. 如果以 $f(\eta)$ 作为 ξ 的预测，其中 f 为波雷尔函数，那么，使预测之均方误差 $E[\xi - f(\eta)]^2$ 达到最小的 f 应是 $f(\eta) = E(\xi|\eta)$.

证明: 我们以连续型为例证明这一定理. 假定 (ξ,η) 有联合密度函数 $p(x,y)$ 及边缘密度 $p_1(x)$ 与 $p_2(x)$, 利用全期望公式 (2.17) 可知，目标函数

$$
\begin{aligned}
E[\xi - f(\eta)]^2 &= E\{E[\xi - f(\eta)]^2|\eta\} \\
&= \int_{-\infty}^{+\infty} E\left([\xi - f(y)]^2|\eta = y\right) p_2(y)\, dy.
\end{aligned}
$$

回忆 3.2.2 节的定理 2.3 及其后的讨论，$E([\xi - f(y)]^2|\eta = y)$ 是已知 $\eta = y$ 条件下 ξ 的条件分布关于直线 $x = f(y)$ 为轴的转动惯量，由方差的性质 4° 知，当且仅当 $f(y) = E(\xi|\eta = y)$ 时它达到其最小值 $D(\xi|\eta = y)$. 再因 $p_2(y)$ 非负，便得证 $f(\eta) = E(\xi|\eta)$ 使 $E[\xi - f(\eta)]^2$ 达最小. 定理得证.

于是，当 (ξ,η) 为例 2.8 中随机向量时，$\xi/2$ 是对 η 的最优预测. 当 (ξ,η) 为二元正态随机向量时，由 (2.16) 式知

$$
f(\eta) = a_1 + r\frac{\sigma_2}{\sigma_1}(\eta - a_2)
$$

是对分量 ξ 的最优预测. 注意，正态情形的最优预测是线性预测. 就是说，从整个波雷尔函数类中选择 f 使均方误差 $E[\xi - f(\eta)]^2$ 达到最小，等价于从线性函数类中所选出的最优预测，这一性质在概率论的应用中十分重要.

3.2.6 矩

以上运用随机变量的一，二阶矩，给出了数学期望，方差与协方差这三种最重要的数字特征．如果需要进一步研究更精细的数字特征，则需计算随机变量的高阶矩．

定义 2.6 如果 $E|\xi|^k < +\infty$, 则称 $m_k = E(\xi^k)$ 为随机变量 ξ(及其分布) 的 k 阶原点矩, 而称 $c_k = E[\xi - E(\xi)]^k$ 为随机变量 ξ(及其分布) 的 k 阶中心矩.

显然，数学期望是一阶原点矩 m_1，方差是二阶中心矩 c_2. 运用初等不等式 $|\xi|^k \le 1 + |\xi|^{k+1}$ 知，若随机变量 ξ 的高阶矩有限，则其低阶矩也有限．

直接计算可得知，原点矩与中心矩可以相互表出：

定理 2.11 当 $E|\xi|^k < +\infty$ 时有

$$c_k = \sum_{i=0}^{k} \binom{k}{i} (-m_1)^{k-i} m_i, \tag{2.20}$$

与

$$m_k = \sum_{i=0}^{k} \binom{k}{i} c_{k-i} m_1^i. \tag{2.21}$$

例 2.10 正态分布的原点矩 设 ξ 服从 $N(0,1)$ 分布，则对任何自然数 n 有

$$\int_{-\infty}^{+\infty} |x|^n \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx < +\infty.$$

故当 $n = 2k-1$ 为奇数时 $m_{2k-1} = 0$, 当 $n = 2k$ 为偶数时

$$m_{2k} = \int_{-\infty}^{+\infty} |x|^{2k} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} x^{2k}\, e^{-\frac{x^2}{2}}\, dx$$

$$= \frac{2^k}{\sqrt{\pi}} \int_0^{+\infty} t^{k-\frac{1}{2}} e^{-t} dt = \frac{2^k}{\sqrt{\pi}} \Gamma(k+\frac{1}{2})$$

$$= (2k-1)(2k-3)\cdot\cdots\cdot 3\cdot 1 = (2k-1)!!.$$

3.2.7 习题

1. 求习题 3.1.6 第 1 题中 ξ 的方差.

2. 求习题 3.1.6 第 2 题中 ξ 的方差.

3. 求习题 2.5.5 第 1 题中 ξ,η 的协方差.

4. 求习题 2.5.5 第 6 题中 ξ,η 的协方差.

5. 试求二项分布的方差.

6. 袋中有编号 1 至 n 的 n 张卡片，现从中任取出 m 张. 试对 (1) 有放回；(2) 不放回 $(m \le n)$ 二情形求 m 张卡片上编号之和的方差.

7. 求习题 3.1.6 的第 13 题中戴对自己帽子的人数 ξ 的方差 $(n > 1)$.

8. 设随机变量 (ξ,η,ζ) 有联合密度函数

$$p(x,y,z) = (x+y)z\,e^{-z}, \qquad 0 < x,y < 1,\ z > 0.$$

求此随机变量的协方差阵.

9. 设随机变量 $\xi_1,\cdots,\xi_{m+n}$ $(m < n)$ 是独立同分布的，它们有有限的方差. 求 $\alpha = \xi_1 + \cdots + \xi_n$, 与 $\beta = \xi_{m+1} + \cdots + \xi_{m+n}$ 之间的相关系数.

10. 设 ξ,η 都是只有两个可能值的随机变量. 试证，若 ξ,η 不相关，则 ξ 与 η 独立.

11. 设平方可积随机变量 ξ 的密度函数为偶函数，试证 ξ 与 $|\xi|$ 不相关，但它们不相互独立.

12. 袋中有 N 个球，其中白球数 τ 为随机变量， $E(\tau) = n$. 现从袋中有放回地任取 m 个球，求取出的白球个数 ξ 的数学期望.

13. 设随机向量 (ξ,η) 有联合密度函数

$$p(x,y)=\frac{1}{y}\,e^{-\frac{x}{y}-y},\qquad x,y>0.$$

求 $E(\xi|\eta)$ 及 $D(\xi|\eta)$.

14. 设某矿山在一个月中发生事故数服从普阿松分布，其参数 λ 是有如下分布的随机变量

$$(1)\ \begin{pmatrix}2 & 3\\ 0.6 & 0.4\end{pmatrix};\qquad (2)\ p_\lambda(t)=\ e^{-t},\ \ t>0.$$

求此矿山在一个月中无事故的概率及恰发生 3 次事故的概率.

15. 设随机向量 (ξ,η) 的联合密度函数为

$$p(x,y)=\frac{1}{y}\,e^{-y},\qquad 0<x<y.$$

求 $E(\xi^3|\eta)$.

16. 设随机变量 ξ_1 在区间 $(0,1)$ 上均匀分布. 对 $k\geq 1$, 若已知 $\xi_k=x_k$, 则 ξ_{k+1} 在区间 (x_k,x_k+1) 上均匀分布，求 $E(\xi_n)$.

17. 试证对任何平方可积的随机向量 (ξ,η) 有

$$D(\xi)=E[D(\xi|\eta)]+D[E(\xi|\eta)].$$

18. 试证明本节定理 2.11.

19. 试求参数为 λ 的指数分布的 k 阶原点矩 m_k.

20. 试求 $\chi^2(n)$ 分布的 k 阶原点矩 m_k.

3.3 母函数

3.3.1 定义与例

分布函数及其密度无疑是描述随机变量概率规律的最有力的工具. 尤其是它具备明确的概率含义，故运用分布函数可以方便地解决许多与随机变量有关的概率问题. 正是由于这一点，标准的概率论教科书的大部分内容都是用分布函数来阐述的. 但是，在今后的一些论

题中，分布函数也表现出某种不足. 例如，分布函数本身的分析性质不大好，它只是一个单边连续的有界非降函数. 又如，独立随机变量和的分布函数等于各分布函数的卷积，这在计算上带来不少麻烦. 而独立和问题是近代概率论中一个重要研究课题. 再如，按分布函数分型，离散型与连续型泾渭分明. 但是象正态分布这种最重要的连续型分布，它的极限都可以是退化到一点的离散型分布，此时，无法用分布函数给出统一的表示式. 以上例证说明分布函数不能适合我们的某些需要，而必须引用其它工具作为补充. 我们先对非负整值随机变量建立其母函数.

定义 3.1 对任何实数列 $\{p_n\}$，如果幂级数

$$G(s)=\sum_{n=0}^{\infty}p_n s^n \tag{3.1}$$

的收敛半径 $s_0>0$，则称 $G(s)$ 为数列 $\{p_n\}$ 的**母函数**. 特别当 $\{p_n\}$ 为某非负整值随机变量 ξ 的概率分布时，(3.1) 式至少在区间 $[-1,1]$ 上绝对收敛且一致收敛. 此时有

$$G(s)=E(s^{\xi}),$$

称此 $G(s)$ 为随机变量 ξ 或其概率分布 $\{p_n\}$ 的母函数.

由定义易求常见分布的母函数.

例 3.1 普阿松分布的母函数

$$G(s)=\sum_{k=0}^{\infty}s^k p(k;\lambda)=\sum_{k=0}^{\infty}s^k\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda(s-1)},$$

此级数在 $s\in(-\infty,+\infty)$ 上收敛.

例 3.2 几何分布的母函数

$$G(s)=\sum_{k=1}^{\infty}s^k g(k;p)=\sum_{k=1}^{\infty}s^k q^{k-1}p=\frac{ps}{1-qs},$$

由于此级数当且仅当 $|qs|<1$ 时收敛，故几何分布之母函数的收敛域为区间 $(-1/q,1/q)$.

母函数 $G(s)$ 显然由概率分布通过 (3.1) 式唯一确定. 反过来说, 由于 $G(s)$ 至少在区间 (-1.1) 内可逐项求 n 阶导数, 再令 $s=0$ 便得

$$p_n = \frac{1}{n!}G^{(n)}(0). \tag{3.2}$$

这说明 $\{p_n\}$ 由其母函数 $G(s)$ 按 (3.2) 式唯一确定. 实际上 (3.2) 是 (3.1) 的反演公式. 至此得知非负整值概率分布 $\{p_n\}$ 与其母函数 $G(s)$ 是一一对应的, 故母函数可以作为描述这种分布的一个工具.

回忆以分布函数刻画概率分布时, 期望与方差等数字特征都是通过对分布函数的积分而求出的. 下面的定理表明, 可以通过求母函数 $G(s)$ 在 $s=1$ 处导数来计算整值概率分布的数字特征.

定理 3.1 设非负整值随机变量 ξ 的母函数为 $G(s)$. 如果 $E(\xi)$ 与 $E(\xi^2)$ 有限, 那么

$$G'(1) = E(\xi), \tag{3.3}$$

$$G''(1) = E(\xi^2) - E(\xi). \tag{3.4}$$

证明: 将 (3.1) 式逐项求导得到

$$G'(s) = \sum_{n=0}^{\infty} ns^{n-1}p_n, \qquad G''(s) = \sum_{n=0}^{\infty} n(n-1)s^{n-2}p_n.$$

(3.1) 式右方级数在区间 $(-1,1)$ 内的一致收敛性保证上述求导至少在这个区间内可行, 而 $E(\xi)$ 与 $E(\xi^2)$ 有限则保证求导后的两级数在 $s=1$ 处收敛, 从而其和函数连续. 至此 (3.3) 与 (3.4) 式得证.

3.3.2 独立和的母函数

对于相互独立的非负整值随机变量, 我们已经在 2.7.1 节中讨论过其和的分布, 导出了离散卷积公式. 如下定理表明, 母函数可将卷积化为普通的乘积, 从而简化了计算.

定理 3.2 设 ξ,η 为相互独立的整值随机变量，它们分别有概率分布 $\{a_n\},\{b_n\}$ 及对应的母函数 $A(s),B(s)$. 则它们的和 $\xi+\eta$ 的母函数为

$$C(s)=A(s)\,B(s). \tag{3.5}$$

证明: 由于母函数 $A(s)$ 与 $B(s)$ 在区间 $[-1,1]$ 上绝对收敛且一致收敛，故在此区间内它们的乘积为

$$A(s)\,B(s)=\sum_{i=0}^{\infty}a_is^i\cdot\sum_{j=0}^{\infty}b_js^j=\sum_{n=0}^{\infty}\sum_{k=0}^{n}a_kb_{n-k}s^n.$$

上式的右方是数列 $\left\{\sum\limits_{k=0}^{n}a_kb_{n-k}\right\}$ 的母函数. 但由离散卷积公式知

$$P\{\xi+\eta=n\}=\sum_{k=0}^{n}a_kb_{n-k},\qquad n=0,1,2,\cdots.$$

由母函数的唯一性就证明了 $A(s)\,B(s)$ 是 $\xi+\eta$ 的母函数. 定理证完.

反复运用 (3.5) 式可得，如果 $\xi_1,\cdots,\xi_n$ 相互独立，有共同的母函数 $G(s)$, 那么它们的和 $\sum\limits_{k=1}^{n}\xi_k$ 的母函数为 $[G(s)]^n$. 现应用这个结论求二项分布的母函数.

例 3.3 二项分布的母函数 设 ξ 服从 $b(k;n,p)$ 分布. 周知，ξ 可化为 n 个相互独立的贝努里随机变量之和 $\xi=\sum\limits_{k=1}^{n}\xi_k$, 每个 ξ_k 的母函数 $P(s)=qs^0+ps^1=q+ps$, 故用定理 3.2 得 ξ 的母函数为

$$G(s)=[P(s)]^n=(q+ps)^n,\qquad s\in(-\infty,+\infty).$$

再设 η 是与上述 ξ 独立的随机变量，η 服从 $b(k;m,p)$ 分布，其母函数 $F(s)=(q+ps)^m$. 由定理 3.2 知 $\xi+\eta$ 的母函数为 $G(s)F(s)=(q+ps)^{n+m}$. 用母函数的唯一性得证 $\xi+\eta$ 服从 $b(k;n+m,p)$ 分布. 至此，我们再次证得 2.7.1 节例 7.1 中的结论.

下面介绍一个运用母函数求概率的例题.

例 3.4 掷均匀骰子 5 枚，求点数和为 15 的概率.

解: 以 ξ_i 表示第 i 枚骰子掷得的点数，则 $\xi_1,\cdots,\xi_5$ 为相互独立的随机变量，它们都服从 1 至 6 这 6 个数上的等可能分布. 因此每个 ξ_i 的母函数均为

$$G(s)=\frac{1}{6}(s+s^2+s^3+s^4+s^5+s^6),\qquad s\in(-\infty,+\infty).$$

用定理 3.2 知，点数和 ξ 的母函数为 $[G(s)]^5$，于是所求概率恰为 $[G(s)]^5$ 展开式中 s^{15} 项的系数 p_{15}，注意到

$$\begin{aligned}[G(s)]^5&=\frac{s^5}{6^5}(1+s+s^2+s^3+s^4+s^5)^5=\frac{s^5}{6^5}(1-s^6)^5(1-s)^{-5}\\&=\frac{s^5}{6^5}(1-5s^6+10s^{12}-10s^{18}+5s^{24}-s^{30})\cdot\\&\qquad(1+\cdots+70s^4+\cdots+1001s^{10}+\cdots),\end{aligned}$$

故所求概率为

$$p_{15}=\frac{1}{6^5}(1001-5\times 70)=\frac{651}{7776}.$$

下面考虑随机多个非负整值随机变量之和的母函数.

定理 3.3 设 $\{\xi_k\}$ 为相互独立的非负整值随机变量序列，它们有共同的母函数 $G(s)$. 如果 η 是另一非负整值随机变量，其母函数为 $F(s)$. 那么当 η 与每个 ξ_k 均独立时，$\xi=\sum\limits_{k=1}^{\eta}\xi_k$ 的母函数为

$$H(s)=F[G(s)]. \tag{3.6}$$

证明: 用全期望公式及 (3.5) 式立得 ξ 的母函数

$$\begin{aligned}H(s)&=E(s^{\xi})=E[E(s^{\xi}|\eta)]\\&=\sum_{n=0}^{\infty}P\{\eta=n\}E(s^{\xi}|\eta=n)\\&=\sum_{n=0}^{\infty}P\{\eta=n\}[G(s)]^n=F[G(s)].\end{aligned}$$

定理证毕.

如果所涉及的随机变量均可积，将 (3.6) 式对 s 求导可得

$$H'(s) = F'[G(s)]\,G'(s). \tag{3.7}$$

令 $s=1$, 注意到 $G(1)=1$ 并运用 (3.3) 式可得

$$E(\xi) = E(\eta)E(\xi_1).$$

于是对非负整值分布情形证明上节的 (2.18) 式. 将 (3.7) 式再次对 s 求导数得

$$H''(s) = F'[G(s)]G''(s) + F''[G(s)][G'(s)]^2.$$

令 $s=1$, 运用 (3.3) 及 (3.4) 式可推出

$$E(\xi^2) - E(\xi) = E(\eta)E(\xi_1^2) - E(\eta)E(\xi_1)$$

$$+E(\eta^2)[E(\xi_1)]^2 - E(\eta)[E(\xi_1)]^2,$$

于是有

$$D(\xi) = E(\eta)D(\xi_1) + D(\eta)[E(\xi_1)]^2.$$

这就是上节 (2.19) 式. 我们以母函数为工具，得到非负整值随机变量情形的简捷证明.

3.3.3 习题

1. 求下列概率分布 $\{p_k\}$ 对应的母函数:

(1) $p_k = \dfrac{\lambda^k e^{-\lambda}}{k!(1-e^{-\lambda})}, \qquad k=1,2,\cdots;$

(2) $p_k = pq^k(1-q^{N+1}), \qquad k=0,1,\cdots,N.$

2. 求巴斯卡分布 $f(k;r,p)$ 的母函数，并用母函数求其期望与方差.

3. 设非负整值随机变量 ξ 的母函数为 $G(s)$, 对非负整数 m,n 求 $m\xi+n$ 的母函数.

4. 某城镇共有 1000 辆汽车，牌照号自 000 至 999. 试用母函数求在此城街上任遇一汽车，其牌照号数字之和等于 9 的概率.

5. 假定某公共汽车站在 $[0,t)$ 内来到的乘客批数 η 服从参数为 λt 的普阿松分布 $(\lambda>0)$. 如果各批来到顾客人数相互独立，且来 n 个乘客的概率均为 $p_n\ (n=1,2,\cdots)$, 试求 $[0,t)$ 时段内来到汽车站乘客数 ξ 的母函数与数学期望.

6. 假定非负整值随机变量 ξ 的母函数为 $G(s)$. 试求数列
(1) $a_n=P\{\xi\le n\}$;　　(2) $b_n=P\{\xi=2n\}$ 对应的母函数.

7. 甲与乙两人各抛掷均匀硬币 n 次. 试利用母函数计算，甲所得正面的个数比乙的正面个数多 $k\ (0\le k\le n)$ 个的概率.

8. 在可列重贝努里试验中，以 μ_n 表示前 n 次试验中成功偶数次的概率，求 $\{\mu_n\}$ 的母函数.

9. 将成功率为 p 的试验独立重复 τ 次，这里 τ 是与每次成功与否均独立的非负整值随机变量. 试用母函数法证明，如果其中成功与失败的次数 ξ 与 η 相互独立，则 τ 必服从普阿松分布.（参见 2.6.2 节的例 6.1 及其后的讨论.）

3.4 特征函数

3.4.1 定义和基本性质

我们已经简要地介绍了母函数, 并运用它们研究了整值概率分布的某些性质. 实际上，这只不过是引入描述概率分布另一重要工具特征函数的先导. 特征函数是分布函数的富里埃变换，它与分布函数一一对应. 虽然特征函数不象分布函数那样具有直观的概率含义，但它却有极好的分析性质. 因此在处理某些概率问题，特别是在独立随机变量和的分布问题中起着非常重要的作用.

设 ξ,η 是概率空间 $(\Omega,\mathcal{F},P)$ 中两个实值随机变量, 由 $\zeta=\xi+i\eta$ (其中 i 为虚单位， $i^2=-1$) 定义了一个复值随机变量 ζ, 可以把 ζ

作为实值随机向量 (ξ,η) 来处理. 例如, ζ 的数学期望可按线性定义为

$$E(\zeta)=E(\xi)+iE(\eta).$$

又如, $\zeta_1=\xi_1+i\eta_1$ 与 $\zeta_2=\xi_2+i\eta_2$ 相互独立, 当且仅当随机向量 (ξ_1,η_1) 与 (ξ_2,η_2) 独立, 当且仅当对任何波雷尔函数 f,g 均有

$$E[f(\zeta_1)g(\zeta_1)]=E[f(\zeta_1)]\,E[g(\zeta_2)]. \tag{4.1}$$

(4.1) 式实际上是 3.1.2 节中 (1.16) 式的二维形式.

定义 4.1 设 $F(x)$ 为 $\Re=(-\infty,+\infty)$ 上的一个分布函数, 称

$$f(t)=\int_{-\infty}^{+\infty} e^{itx}\,dF(x),\qquad t\in\Re \tag{4.2}$$

为 $F(x)$ 的**特征函数**. 如果 $F(x)$ 是随机变量 ξ 的分布函数, 则此 $f(t)$ 也称为 ξ 的特征函数, 此时有

$$f(t)=E(\,e^{it\xi}).$$

注意对一切 $t,x\in\Re$ 有 $|\,e^{itx}|=1$, 故 (4.2) 式右方积分的模不超过 1. 这就是说, 对于无论怎样的概率分布, 它的特征函数总是唯一确定地存在着, 而不象母函数那样只对一部分概率分布才存在. 特征函数的这种普遍性, 是它在近代概率论中有广泛应用的原因之一.

对离散型与连续型这两种常见的概率分布, (4.2) 式分别写作

$$f(t)=\sum_k e^{itx_k}p_k,\qquad \text{其中}\ \ p_k=P\{\xi=x_k\}. \tag{4.3}$$

与

$$f(t)=\int_{-\infty}^{+\infty} e^{itx}p(x)\,dx,\ \ \text{其中}\ \ p(x)\ \text{为密度函数}. \tag{4.4}$$

在计算特征函数时, 经常会运用欧拉公式

$$e^{it}=\cos t+i\sin t\,,$$

于是 (4.4) 式化为

$$f(t)=\int_{-\infty}^{+\infty}\cos tx\,p(x)\,dx+i\int_{-\infty}^{+\infty}\sin tx\,p(x)\,dx. \tag{4.5}$$

例 4.1 离散型分布的特征函数 运用 (4.3) 式立得，贝努里分布的特征函数

$$f(t)=q+p\,e^{it};$$

普阿松分布的特征函数

$$f(t)=\sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda}= e^{\lambda(e^{it}-1)};$$

而几何分布的特征函数为

$$f(t)=\sum_{k=1}^{\infty} e^{itk}q^{k-1}p=\frac{p\,e^{it}}{1-q\,e^{it}}.$$

上述结果表明，如果整值随机变量的母函数为 $G(s)$，那么它的特征函数 $f(t)=G(e^{it})$.

例 4.2 指数分布的特征函数 用 (4.5) 式可知

$$\begin{aligned}f(t)&=\int_0^{+\infty}\lambda\,e^{-\lambda x}\cos tx\,dx+i\int_0^{+\infty}\lambda\,e^{-\lambda x}\sin tx\,dx\\&=\lambda(J_1+iJ_2).\end{aligned}$$

用分部积分便得，其中

$$J_1=\frac{\lambda}{t}J_2 \qquad 而 \qquad J_2=\frac{1}{t}-\frac{\lambda}{t}J_1.$$

由此可解出

$$J_1=\frac{\lambda}{\lambda^2+t^2} \qquad 与 \qquad J_2=\frac{t}{\lambda^2+t^2}.$$

代回便得指数分布的特征函数为

$$f(t)=\frac{\lambda(\lambda+it)}{\lambda^2+t^2}=\frac{\lambda}{\lambda-it}=\left(1-i\frac{t}{\lambda}\right)^{-1}.$$

下面定理包含着特征函数的基本性质.

定理 4.1 设 $f(t)$ 是由 (4.2) 式定义的特征函数, 则有

1° $|f(t)| \le f(0) = 1$;

2° 共轭对称性: $f(-t) = \overline{f(t)}$;

3° $f(t)$在 $t \in (-\infty, +\infty)$ 上一致连续;

4° 半正定性: 任意 $n \ge 1$, 任意 n 个实数 $t_1, \cdots, t_n$ 及 n 个复数 $a_1, \cdots, a_n$ 有

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j \overline{a}_k f(t_j - t_k) \ge 0; \tag{4.6}$$

5° $f_{a+b\xi}(t) = e^{iat} f_\xi(bt)$.

证明: 直接计算立得 1° 与 2°. 考虑 t 有改变量 Δt, 相应的 $f(t)$ 增量的模

$$\begin{aligned} |f(t+\Delta t) - f(t)| &\le \int_{-\infty}^{+\infty} \left| e^{i(t+\Delta t)x} - e^{itx} \right| dF(x) \\ &= \int_{-\infty}^{+\infty} \left| e^{i(t+\frac{\Delta t}{2})x} \right| \left| e^{i\frac{\Delta t}{2}x} - e^{-i\frac{\Delta t}{2}x} \right| dF(x) \\ &= \int_{-\infty}^{+\infty} 2 \left| \sin \frac{\Delta t}{2} x \right| dF(x). \end{aligned}$$

上式右方已不含 t, 且当 $\Delta t \to 0$ 时可以任意地小. 至此得 $f(t)$ 的一致连续性 3°. 再者, (4.6) 式的左端

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j \overline{a}_k f(t_j - t_k) = E \left| \sum_{j=1}^{n} a_j \ e^{it_j \xi} \right|^2,$$

运用期望的序性立得证 $f(t)$ 的半正定性. 最后, 对任何常数 a, b, 随机变量 $a + b\xi$ 的特征函数

$$\begin{aligned} f_{a+b\xi}(t) &= E\left[e^{it(a+b\xi)} \right] = e^{iat} E\left[e^{i(tb)\xi} \right] \\ &= e^{iat} f_\xi(bt), \end{aligned}$$

这就是性质 5°. 定理证毕.

例 4.3 均匀分布的特征函数 设 η 服从区间 (a,b) 内的均匀分布. 由于它与区间 (-1.1) 上均匀分布随机变量 ξ 有线性关系

$$\eta = \frac{a+b}{2} + \frac{b-a}{2}\xi,$$

故用定理 4.1 中 5° 知, 可先求 ξ 的特征函数. 由 (4.5) 式可直接算出 ξ 的特征函数

$$f_\xi(t) = \frac{1}{2}\int_{-1}^{1} \cos tx \, dx + \frac{i}{2}\int_{-1}^{1} \sin tx \, dx = \frac{\sin t}{t}.$$

这里 (及今后) 需在函数的可去间断点 $t=0$ 处按连续性补设 $f_\xi(0) = 1$. 再用性质 5° 便得区间 (a,b) 上均匀分布的特征函数

$$f_\eta(t) = e^{i\frac{a+b}{2}t} f_\xi\left(\frac{b-a}{2}t\right) = \frac{e^{ibt} - e^{iat}}{it(b-a)}.$$

在进一步研究特征函数的性质之前, 需要做一些准备.

引理 4.2 对一切 $x \in \Re$ 及非负整数 n 有

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| \le \left[\frac{|x|^{n+1}}{(n+1)!}\right] \wedge \left[\frac{2|x|^n}{n!}\right]. \tag{4.7}$$

证明: 用分部积分可得

$$\int_0^x (x-s)^n \, e^{is} \, ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1} \, e^{is} \, ds. \tag{4.8}$$

递归可得知, 对一切 $n \ge 0$ 有

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n \, e^{is} \, ds. \tag{4.9}$$

由此推出估计式

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| \le \frac{|x|^{n+1}}{(n+1)!}. \tag{4.10}$$

将 (4.8) 式中 n 换为 $n-1$, 解出右方积分并代入 (4.9) 便推出

$$e^{ix}=\sum_{k=0}^{n}\frac{(ix)^k}{k!}+\frac{i^n}{(n-1)!}\int_0^x(x-s)^{n-1}(e^{is}-1)\,ds.$$

由此可得另一估计式

$$\left|e^{ix}-\sum_{k=0}^{n}\frac{(ix)^k}{k!}\right|\le\frac{2|x|^n}{n!}. \tag{4.11}$$

联合 (4.10) 与 (4.11) 便得到 (4.7) 式，引理证完.

这个引理给出指数函数泰勒公式的余项估计，对于较小的 $|x|$, 我们用 (4.7) 式右端的第一项作为上界，而当 $|x|$ 较大时，则 (4.7) 式右方的第二项更为有效. 我们写出此估计当 $n=0,1,2$ 时的特例以备引用：

$$|e^{ix}-1|\le|x|\wedge 2; \tag{4.12}$$

$$|e^{ix}-1-ix|\le\frac{x^2}{2}\wedge(2|x|); \tag{4.13}$$

$$\left|e^{ix}-1-ix+\frac{x^2}{2}\right|\le\left(\frac{1}{6}|x|^3\right)\wedge x^2. \tag{4.14}$$

易见，对一切 $x\in\Re$, 当 $n\to\infty$ 时 (4.7) 式的右方趋于 0, 从而得到熟知的指数函数的泰勒展式

$$e^{itx}=\sum_{k=0}^{\infty}\frac{(itx)^k}{k!},\qquad t,x\in\Re. \tag{4.15}$$

进一步地，如果此式右方可逐项对分布函数 $F(x)$ 取积分，且积分后的级数收敛到特征函数 $f(t)$, 我们就得到 $f(t)$ 的幂级数展式

$$f(t)=\sum_{k=0}^{\infty}\frac{(it)^k}{k!}E(\xi^k). \tag{4.16}$$

当然，此式成立的一个必要条件是 ξ 的各阶矩 $E(\xi^k)$ 有限，再者就是要求余项的模趋于 0.

定理 4.3 如果随机变量 ξ 的各阶原点矩有限，那么对一切满足

$$\lim_n \frac{|t|^n E|\xi|^n}{n!} = 0 \tag{4.17}$$

的 t, ξ 的特征函数 $f(t)$ 有展开式 (4.16) 成立.

证明: 运用估计式 (4.10) 立得，在 (4.17) 式条件下有

$$\left| f(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!} E(\xi^k) \right| \leq \frac{|t|^{n+1} E|\xi|^{n+1}}{(n+1)!} \to 0.$$

定理于是得证.

可以用展开式 (4.16) 计算某些符合条件的概率分布的特征函数.

例 4.4 正态分布的特征函数 先考虑标准正态分布的随机变量 ξ. 由 3.2.6 节的例 2.10 知它的各阶矩存在，且对一切 $r = 1, 2, \cdots$ 有

$$E(\xi^{2r}) = (2r-1)!!, \qquad E(\xi^{2r-1}) = 0. \tag{4.18}$$

从而对一切 $t \in \Re$ 有

$$\lim_r \frac{|t|^{2r} E|\xi|^{2r}}{(2r)!} = \lim_r \frac{|t|^{2r}}{(2r)!!} = \lim_r \left|\frac{t^2}{2}\right|^r \frac{1}{r!} = 0,$$

即 $n = 2r$ 时 (4.17) 式成立. 另一方面，对 $r \geq 1$, 有

$$E|\xi|^{2r+1} = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} x^{2r+1}\, e^{-\frac{x^2}{2}}\, dx = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} 2^r t^r\, e^{-t}\, dt$$

$$= \sqrt{\frac{2}{\pi}} 2^r \Gamma(r+1) = \sqrt{\frac{2}{\pi}} (2r)!!.$$

故

$$\lim_r \frac{|t|^{2r+1} E|\xi|^{2r+1}}{(2r+1)!} = \sqrt{\frac{2}{\pi}} \lim_r \frac{|t|^{2r+1}}{(2r-1)!!}$$

$$\leq \sqrt{\frac{2}{\pi}} |t| \lim_r \frac{|t|^{2r}}{(2r)!!} = 0.$$

总之，对一切 $t \in \Re$ 有 (4.17) 式成立. 将 (4.18) 代入 (4.16) 便得到 ξ 的特征函数

$$f(t) = \sum_{r=0}^{\infty} \frac{(it)^{2r}}{(2r)!}(2r-1)!! = \sum_{r=0}^{\infty} \frac{1}{r!}\left(\frac{-t^2}{2}\right)^r = e^{-\frac{t^2}{2}}.$$

最后用定理 4.1 中性质 5° 便得一般正态分布 $N(a,\sigma^2)$ 的特征函数为

$$f_{a+\sigma\xi}(t) = e^{iat-\frac{1}{2}\sigma^2 t^2}.$$

在定理 4.3 的条件下，由 (4.16) 式及泰勒级数的唯一性立得特征函数在 $t=0$ 处各阶导数存在，且有

$$f^{(k)}(0) = i^k E(\xi^k), \qquad k \geq 1. \tag{4.19}$$

但下面定理说明，为导出 (4.19) 我们不需要条件 (4.17).

定理 4.4 设随机变量 ξ 的 k 阶原点矩有限，则其特征函数 k 阶可微，且有

$$f^{(k)}(t) = E\left[(i\xi)^k\, e^{it\xi}\right]. \tag{4.20}$$

证明: 先考虑 1 阶导数，我们有

$$\frac{1}{\Delta t}[f(t+\Delta t - f(t)] - E[i\xi\, e^{it\xi}]$$

$$= E\left[e^{it\xi}\cdot\frac{1}{\Delta t}\left(e^{i\cdot\Delta t\cdot\xi} - 1 - i\cdot\Delta t\cdot\xi\right)\right].$$

运用 (4.13) 式，一方面上式右端方括号的模不超过 $2|\xi|$, 后者可积；另一方面上式右端方括号的模不超过 $\frac{1}{2}\cdot\Delta t\cdot\xi^2$. 令 $\Delta t \to 0$, 用勒贝格控制收敛定理得证 $f'(t)$ 存在且有

$$f'(t) = E(i\xi\, e^{it\xi}).$$

归纳地重复上述论证便得证 (4.20) 式. 定理得证.

注意此定理的逆不成立，即特征函数可导不能推出相应的矩存在.

下面考虑独立随机变量和的特征函数.

定理 4.5 设随机变量 ξ_1 与 ξ_2 相互独立，则它们之和的特征函数等于各自特征函数之积，即有

$$f_{\xi_1+\xi_2}(t) = f_{\xi_1}(t)\, f_{\xi_2}(t).$$

证明: 运用定义及 (4.1) 式立即得证

$$f_{\xi_1+\xi_2}(t) = E(\, e^{it\xi_1}\, e^{it\xi_2}) = E(\, e^{it\xi_1})E(\, e^{it\xi_2}) = f_{\xi_1}(t)f_{\xi_2}(t).$$

此定理把分布函数的卷积运算化为相应之特征函数的普通乘积，从而大大简化了计算．易见，对于任意有限个独立随机变量的情形，结论仍然成立．

例 4.5 二项分布的特征函数 熟知，二项分布随机变量是 n 个相互独立的贝努里随机变量之和，运用定理 4.5 及例 4.1 中结果可知，二项分布的特征函数为 $f(t) = (q + p\, e^{it})^n$.

3.4.2 反演公式与唯一性定理

由定义 4.1 及其后的讨论可知，特征函数 $f(t)$ 由其相应的分布函数按 (4.2) 式唯一确定．现在的问题是，特征函数是否也唯一地确定分布函数？会不会有不同的分布函数，通过 (4.2) 式算出的特征函数相同呢？本节将回答这个问题，严格地建立起分布函数与特征函数之间的一一对应关系．为此必须建立 (4.2) 式的反演公式，即富里埃逆变换公式．我们先证明一个分析结果．

引理 4.6

$$\lim_{c\to+\infty}\int_0^c \frac{\sin ax}{x}\, dx = \frac{\pi}{2}\mathrm{sgn}\{a\}.$$

其中 $\mathrm{sgn}\{a\}$ 是参数 a 的符号函数．

证明: 记

$$I(a,c) = \int_0^c \frac{\sin ax}{x}\, dx,$$

作积分换元 $y = ax$, 并注意被积函数是偶函数便得

$$\lim_{c\to+\infty} I(a,c) = \mathrm{sgn}\{a\} \lim_{c\to+\infty} I(1,c).$$

于是只需证明

$$\lim_{c\to+\infty} I(1,c) = \lim_{c\to+\infty} \int_0^c \frac{\sin x}{x}\,dx = \frac{\pi}{2}.$$

将

$$\frac{1}{x} = \int_0^{+\infty} e^{-ux}\,du$$

代入 $I(1,c)$ 并交换积分次序得

$$I(1,c) = \int_0^{+\infty} \left[\int_0^c e^{-ux} \sin x\,dx\right] du$$

$$= \int_0^{+\infty} \left[\frac{1}{1+u^2} - \frac{1}{1+u^2}\, e^{-uc}(u\sin c + \cos c)\right] du$$

$$= \frac{\pi}{2} - \int_0^{+\infty} e^{-s}\frac{s\sin c + c\cos c}{s^2+c^2}\,ds$$

当 $c \to +\infty$ 时, 最后一个积分的被积函数趋于 0 且被可积的 e^{-s} 所控制, 用勒贝格控制收敛定理知它趋于 0. 引理于是得证.

定理 4.7 反演公式 设 $f(t)$ 是由 (4.2) 式定义的特征函数, 则对任意的 $a,b \in \Re$ 有

$$\frac{1}{2}[F(b+0)+F(b)] - \frac{1}{2}[F(a+0)+F(a)]$$

$$= \lim_{c\to+\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} f(t)\,dt. \tag{4.21}$$

证明: 当 $a = b$ 时上式左右均为 0, 故不妨设 $a < b$. 考察积分

$$J(c) = \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} f(t)\,dt$$

$$= \frac{1}{2\pi}\int_{-c}^{c}\left[\int_{-\infty}^{+\infty}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\,dF(x)\right]dt.$$

用 (4.12) 式知上述被积函数的模

$$\left|\frac{e^{-iat}-e^{-ibt}}{it}e^{itx}\right| = \frac{\left|e^{it(b-a)}-1\right|}{|t|} \le b-a,$$

从而积分 $J(c)$ 有限，交换积分次序可得

$$J(c) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\int_{-c}^{c}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\,dt\right]dF(x)$$

$$= \frac{1}{\pi}\int_{-\infty}^{+\infty}\left[\int_{0}^{c}\frac{\sin t(x-a)}{t}\,dt - \int_{0}^{c}\frac{\sin t(x-b)}{t}\,dt\right]dF(x)$$

$$= \frac{1}{\pi}\int_{-\infty}^{+\infty}[\,I(x-a,c)-I(x-b,c)]\,dF(x).$$

据引理 4.6 知

$$\lim_{c\to+\infty}\frac{1}{\pi}[\,I(x-a,c)-I(x-b,c)\,] = \begin{cases} 1, & \text{若 } a<x<b, \\ \frac{1}{2}, & \text{若 } x=a \text{ 或 } b, \\ 0, & \text{其它 } x. \end{cases}$$

可见对一切 x 及较大的 c, 此函数有界. 将它在 $(-\infty,+\infty)$ 上对 $F(x)$ 的积分 $J(c)$ 分割为 5 项，并用勒贝格控制收敛定理便得到

$$\lim_{c\to+\infty}J(c) = \int_{(-\infty,a)}0\,dF(x) + \int_{\{a\}}\frac{1}{2}\,dF(x) + \int_{(a,b)}1\,dF(x)$$

$$+\int_{\{b\}}\frac{1}{2}\,dF(x) + \int_{(b,+\infty)}0\,dF(x)$$

$$= \frac{1}{2}[F(b+0)+F(b)] - \frac{1}{2}[F(a+0)+F(a)].$$

定理得证.

定理 4.8 唯一性定理 分布函数由其特征函数唯一确定.

证明: 以 C_F 表分布函数 $F(x)$ 的连续点集, $f(t)$ 为 $F(x)$ 的特征函数. 则当 $a, b \in C_F$ 时, 反演公式 (4.21) 变为

$$F(b) - F(a) = \lim_{c \to +\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} f(t)\, dt.$$

上式先令 a 在 C_F 中趋于 $-\infty$, 便得 $F(x)$ 在每个 $b \in C_F$ 上被 $f(t)$ 唯一确定; 再令 b 在 C_F 中上升趋于 x, 则由分布函数的左连续性得证 $F(x)$ 在每个 $x \in \Re$ 上被 $f(t)$ 唯一确定: 任 $x \in \Re$

$$F(x) = \lim_{\substack{b \uparrow x \\ b \in C_F}} \lim_{\substack{a \to -\infty \\ a \in C_F}} \lim_{c \to +\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} f(t)\, dt. \tag{4.22}$$

定理得证.

至此, 我们通过定义 (4.2) 式与反演公式 (4.21) 式建立了随机变量的分布函数与特征函数间的一一对应. 这种无一例外的对应, 使特征函数成为描述概率分布的又一工具, 成为分布函数的有力补充.

下面用特征函数给出连续型分布的一个充分条件.

定理 4.9 如果特征函数绝对可积, 即

$$\int_{-\infty}^{+\infty} |f(t)|\, dt < +\infty, \tag{4.23}$$

则对应的分布函数 $F(x)$ 为连续型, 且其密度函数为

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} f(t)\, dt. \tag{4.24}$$

证明: 对任意的 a 取 $b_n \in C_F$ 且 $b_n \downarrow a$. 对 a 及 b_n 用反演公式 (4.21) 及 (4.12) 式可得

$$F(b_n) - \frac{1}{2}[F(a+0) + F(a)] \le \frac{b_n - a}{2\pi} \int_{-\infty}^{+\infty} |f(t)|\, dt.$$

故由条件 (4.23) 知 $n \to \infty$ 时上式右方趋于 0, 从而左方 $F(a+0) = F(a)$, 即 $F(x)$ 连续. 再对任意的 x 及 Δx 用反演公式得

$$\frac{F(x+\Delta x) - F(x)}{\Delta x} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-itx} - e^{-it(x+\Delta x)}}{it\Delta x} f(t)\, dt.$$

令 $\Delta x\to 0$, 一方面上式右方被积函数趋于 $e^{-itx}f(t)$, 另一方面, 用 (4.12) 式知被积函数被可积的 $|f(t)|$ 所控制, 于是用勒贝格控制收敛定理得证 (4.24) 式, 定理证毕.

3.4.3 初步应用

我们先运用特征函数证明 n 个相互独立的标准正态随机变量之平方和服从 n 个自由度的 χ^2 分布.

例 4.6 假定 $\xi_1,\cdots,\xi_n$ 为相互独立的随机变量, 都服从 $N(0,1)$ 分布, 证明 $\eta=\xi_1^2+\cdots+\xi_n^2$ 有密度函数

$$p_n(x)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}\,e^{-\frac{x}{2}},\qquad x>0. \tag{4.25}$$

解: 直接用定义 (4.4) 式算 $p_n(x)$ 对应的特征函数

$$f_n(t)=\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\lim_{a\to+\infty}I(a),$$

其中

$$\begin{aligned}I(a)&=\int_0^a x^{\frac{n}{2}-1}\,e^{-(\frac{1}{2}-it)x}\,dx\\&=\left(\frac{1}{2}-it\right)^{-\frac{n}{2}}\int_0^a\left[\left(\frac{1}{2}-it\right)x\right]^{\frac{n}{2}-1}e^{-(\frac{1}{2}-it)x}d\left[\left(\frac{1}{2}-it\right)x\right]\\&=\left(\frac{1}{2}-it\right)^{-\frac{n}{2}}\int_{OA}z^{\frac{n}{2}-1}e^{-z}\,dz,\end{aligned}$$

这里复变量 z 的积分路径 OA 是从原点 O 到点 $A=a/2-iat$ 的直线段 (见图 4.1). 由于解析函数在闭路径上积分为 0, 故有

$$\int_{OA}z^{\frac{n}{2}-1}e^{-z}\,dz=\int_{OB}z^{\frac{n}{2}-1}e^{-z}\,dz+\int_{\overset{\frown}{BA}}z^{\frac{n}{2}-1}e^{-z}\,dz.$$

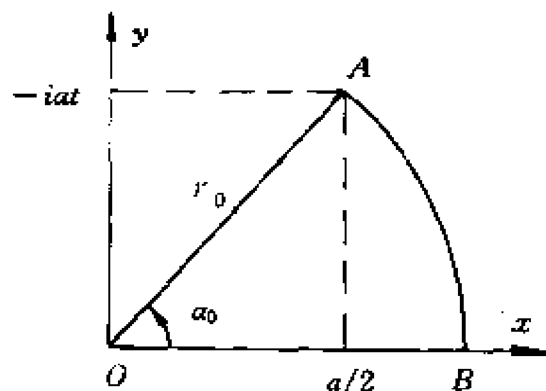


图 4.1 $I(a)$ 的积分路径 $(t<0)$

易见，当 $a \to +\infty$ 时，B 点的坐标 $r_0 = a\sqrt{1+4t^2}/2 \to +\infty$，故

$$\int_{OB} z^{\frac{n}{2}-1}\,e^{-z}\,dz = \int_0^{r_0} x^{\frac{n}{2}-1}\,e^{-x}\,dx \to \Gamma\left(\frac{n}{2}\right).$$

而当 z 沿圆弧 $\overset{\frown}{BA}$ 运动时，其模恒为 r_0；辐角则由 0 变到 $\alpha_0 = \mathrm{tg}^{-1}(-2t)$，故当 $a \to +\infty$ 亦即 $r_0 \to +\infty$ 时，第二个积分的模

$$\left|\int_{BA} z^{\frac{n}{2}-1}\,e^{-z}\,dz\right| \le r_0^{\frac{n}{2}-1}\int_0^{\alpha_0}\left|\,e^{-r_0\cos\theta}\,e^{-ir_0\sin\theta}\right|\,d\theta$$

$$\le r_0^{\frac{n}{2}-1}\alpha_0\,e^{-r_0\cos\alpha_0} \to 0.$$

至此算得由 (4.25) 式中 $p_n(x)$ 定义的 χ^2 分布的特征函数

$$f_n(t) = (1-2it)^{-\frac{n}{2}}.$$

另一方面，在 2.7.2 节的例 7.3 中已经证明，一个标准正态随机变量的平方有密度函数 $p_1(x)$，从而其特征函数为 $f_1(t) = (1-2it)^{-\frac{1}{2}}$. 现 $\xi_1^2,\cdots,\xi_n^2$ 相互独立，有共同的特征函数 $f_1(t)$，由定理 4.5 知 $\eta = \xi_1^2+\cdots+\xi_n^2$ 有特征函数 $(1-2it)^{-\frac{n}{2}} = f_n(t)$. 再由唯一性定理得证 η 的密度函数是 (4.25) 式中的 $p_n(x)$，即 n 个独立的标准正态随机变量之平方和服从 n 个自由度的 χ^2 分布.

到目前为止，我们在计算连续型分布的特征函数例题中，采用过三种方法. 一是运用 (4.5) 式采用纯分析的方法分别计算积分

$$u(t) = \int_{-\infty}^{+\infty}\cos tx\,p(x)\,dx \qquad 与 \qquad v(t) = \int_{-\infty}^{+\infty}\sin tx\,p(x)\,dx.$$

例 4.2 与例 4.3 就是采用的这种方法. 二是当 k 阶绝对矩相当小，使得 (4.17) 式成立时，用定理 4.3 通过幂级数求和式算出特征函数. 在例 4.4 中我们运用这个方法求得正态分布的特征函数. 三是利用复变函数积分的性质，选择适当的积分路径而求出特征函数. 例 4.6 中就是用这种方法算出 χ^2 分布的特征函数. 读者可选用这三种方法中的任何一种. 此外，尚有一种不太严格的方法，即在形式上把 i 视同实

常数算出积分. 例如在例 4.6 中, 直接做变量替换 $y=(1/2-it)x$, 可得

$$f_n(t)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^{+\infty}x^{\frac{n}{2}-1}\,e^{-(\frac{1}{2}-it)x}\,dx$$

$$=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\frac{1}{(\frac{1}{2}-it)^{\frac{n}{2}}}\int_0^{+\infty}y^{\frac{n}{2}-1}\,e^{-y}\,dy=(1-2it)^{-\frac{n}{2}}.$$

由于系数 $(1/2-it)$ 有正实部, 所用变量替换方法仍然适用. 这一点可以严格证明, 在此不予详述.

下面用特征函数研究概率分布的再生性质. 在前面用卷积公式或母函数研究独立随机变量和的分布时, 我们发现有的概率分布有这样的性质: 两个具有同一类型分布的卷积仍是这种类型的分布, 且对应的参数等于各随机变量相应参数之和. 例如二项分布 $b(k;\,n,\,p)$ 与 $b(k;\,m,\,p)$ 的卷积为 $b(k;\,n+m,\,p)$ (参见 3.3.2 节的例 3.3). 当然, 也有的分布不具备这种性质, 例如两指数分布的卷积就不再是指数分布 (参见 2.7.3 节的例 7.4).

定义 4.2 设 $F(x;c)$ 为分布函数, 其中 c 为此概率分布的参数. 如果有

$$F(x;c_1)*F(x;c_2)=F(x;c_1+c_2), \tag{4.26}$$

则称此分布关于参数 c **有再生性**.

由定理 4.5 知, (4.26) 式可用相应的特征函数等价地写作

$$f(x;c_1)f(x;c_2)=f(x;c_1+c_2). \tag{4.27}$$

在常见分布中, 普阿松分布, 正态分布及 χ^2 分布等, 它们的特征函数都是各自参数的指数函数, 满足函数方程 (4.27), 所以都具有再生性. 例如, 我们有

$$N(a_1,\sigma_1^2)*N(a_2,\sigma_2^2)=N(a_1+a_2,\sigma_1^2+\sigma_2^2).$$

3.4.4 多元特征函数

对于多元分布函数同样可引入其富里埃变换，即多元特征函数．它有与一元情形相类似性质与应用，下面予以简要叙述．

定义 n 元分布函数 $F(x_1,\cdots,x_n)$ 所对应的 n 元特征函数为

$$f(t_1,\cdots,t_n)=\int\cdots\int_{\Re^n}e^{i(t_1x_1+\cdots+t_nx_n)}dF(x_1,\cdots,x_n). \tag{4.28}$$

如果 $F(x_1,\cdots,x_n)$ 是随机向量 $(\xi_1,\cdots,\xi_n)$ 的联合分布函数，则有

$$f(t_1,\cdots,t_n)=E\left[e^{i(t_1\xi_1+\cdots+t_n\xi_n)}\right], \tag{4.29}$$

此时称 f 为 $(\xi_1,\cdots,\xi_n)$ 的**联合特征函数**．

这是 n 个实变量 $t_1,\cdots,t_n$ 的复值函数，它有性质

$$|f(t_1,\cdots,t_n)|\leq f(0,\cdots,0)=1,$$

$$f(-t_1,\cdots,-t_n)=\overline{f(t_1,\cdots,t_n)},$$

且在 $\Re^n$ 上是一致连续的．再者，若混合矩 $E[\xi_1^{k_1}\cdots\xi_n^{k_n}]$ 有限，则可用 $(\xi_1,\cdots,\xi_n)$ 的联合特征函数在原点的导数求矩，即有公式

$$E(\xi_1^{k_1}\cdots\xi_n^{k_n})=i^{(-\sum\limits_{j=1}^{n}k_j)}\frac{\partial^{k_1+\cdots+k_n}f(0,\cdots,0)}{\partial t_1^{k_1}\cdots\partial t_n^{k_n}}. \tag{4.30}$$

如果随机向量 $(\xi_1,\cdots,\xi_n)$ 取值于 n 维区间

$$\Delta=\{(x_1,\cdots,x_n):a_j\leq x_j<b_j,\ j=1,\cdots,n\}$$

的边界面上的概率为 0, 则有

$$P\{(\xi_1,\cdots,\xi_n)\in\Delta\}=P\{\bigcap_{j=1}^{n}[a_j\leq\xi_j\leq b_j]\}$$

$$= \lim_{\substack{c_j \to +\infty \\ j=1,\cdots,n}} \frac{1}{(2\pi)^n} \int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \cdots \int_{-c_n}^{c_n} \prod_{j=1}^{n} \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \cdot f(t_1, \cdots, t_n)\, dt_1 \cdots dt_n. \tag{4.31}$$

这就是由特征函数表出分布函数的 n 元反演公式. 这个公式确保随机向量的分布函数与特征函数是一一对应的, 从而可用特征函数来描述 n 元概率分布.

由 (4.29) 式可见, 为求任一子向量 $(\xi_{s_1}, \cdots, \xi_{s_k})$ 的 k 维边缘特征函数, 只要将联合特征函数 $f(t_1, \cdots, t_n)$ 中与子向量的指标 $s_1, \cdots, s_k$ 无干的 t_j 取为 0. 例如, 子向量 $(\xi_2, \cdots, \xi_n)$ 的边缘特征函数为 $f(0, t_2, \cdots, t_n)$. 最后, 随机变量 $\xi_1, \cdots, \xi_n$ 相互独立的充分必要条件是他们的联合特征函数等于每个分量 1 维边缘特征函数之积, 即对一切 $t_1, \cdots, t_n$ 有

$$f(t_1, \cdots, t_n) = f_1(t_1) \cdot \cdots \cdot f_n(t_n). \tag{4.32}$$

3.4.5 习题

1. 求巴斯卡分布 $f(k; r, p)$ 的特征函数, 并用特征函数求其期望.

2. 求伽马分布 $\Gamma(\lambda, r)$ 的特征函数, 并用特征函数求其 k 阶原点矩 m_k.

3. 判断下列函数哪些是特征函数, 并指出其相应的概率分布.

(1) $\cos^2 t$, (2) $\cos t - i \sin t$,

(3) $\dfrac{1-t}{1+t^2}$, (4) $\sin t$,

(5) $\dfrac{1}{1+it}$, (6) $\dfrac{1}{2\,e^{-it} - 1}$,

(7) $\dfrac{1}{1+t^2}$, (8) $|\cos t|$.

4. 若对一切 x 有 $F(x) = 1 - F(-x+0)$, 则称分布函数 $F(x)$ 为**对称**的. 求证分布函数对称的充分必要条件是对应的特征函数为实的偶函数.

5. 证明实值特征函数有性质

$$1-f(2t)\leq 4[1-f(t)], \qquad 1+f(2t)\geq 2[f(t)]^2.$$

6. 设分布函数 $F(x)$ 对应的特征函数为 $f(t)$, 试证对任何 x 有

$$F(x+0)-F(x)=\lim_{c\to+\infty}\frac{1}{2c}\int_{-c}^{c}f(t)\,e^{-itx}\,dt.$$

7. 设

$$g(u)=1-|u|, \qquad |u|<1.$$

(1) 求以 $g(x)$ 为密度函数的特征函数;

(2) 用反演公式求以 $g(t)$ 为特征函数的密度函数.

8. 拉普拉斯分布的密度函数为 $p(x)=\frac{1}{2}\,e^{-|x|}$, 试求此分布的特征函数.

9. 柯西分布 $C(\lambda,\mu)$ 的密度函数为 $p(x)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$, 求证它的特征函数为 $f(t)=\ e^{i\mu t-\lambda|t|}$.

10. 设 ξ 服从柯西分布 $C(1,0)$, $\eta=\xi$, 求证 $f_{\xi+\eta}(t)=f_\xi(t)f_\eta(t)$.

11. 设 $\xi_1,\cdots,\xi_n$ 为相互独立的随机变量, 同服从 $C(1,0)$ 分布, 求 $\overline{\xi}=\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 的分布.

12. 试讨论伽马分布 $\Gamma(\lambda,r)$ 对参数的再生性.

13. 设随机变量 ξ 只取有界区间中的值, 试证明其分布由各阶原点矩 m_k 所确定.

14. 设 $f(t)$ 为某概率分布的特征函数, 试证 $e^{f(t)-1}$ 也是特征函数.

15. 试证随机变量 $\xi_1,\cdots,\xi_n$ 相互独立的充分必要条件是它们的联合特征函数等于各自边缘特征函数的乘积.

16. 设随机向量 (ξ,η) 的联合密度函数为

$$p(x,y)=\frac{1}{4}[1+cxy(x^2-y^2)], \ \ |x|<1 \text{ 且 } |y|<1.$$

其中参数 $c\in(-1,1)$. 试求

(1) 联合特征函数 $f(s,t)$;

(2) 边缘特征函数 $f_1(s)$ 与 $f_2(t)$;

(3) $\xi+\eta$ 的特征函数.

3.5 多元正态分布

3.5.1 密度函数与特征函数

本节以特征函数为主要工具，将 2.4.2 及 2.5.4 节中介绍的正态分布推广到多元情形. 我们总采用列向量，比如 n 维随机向量 $\vec{\xi}$ 及其可能值 $\vec{x}$ 分别为

$$\vec{\xi}=\begin{pmatrix}\xi_1\\ \vdots\\ \xi_n\end{pmatrix} \quad 与 \quad \vec{x}=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}.$$

当 $n=2$ 时，我们在第二章已引入二元正态分布. 这是 $\Re^2$ 上的一个连续型分布，其联合密度函数由 2.5.4 节的 (5.21) 式给出，期望向量 $\vec{a}$ 与协方差阵 B 分别为

$$\vec{a}=\begin{pmatrix}a_1\\ a_2\end{pmatrix}, \qquad B=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}.$$

这里，方阵 B 是一个对称正定阵，其逆阵为

$$B^{-1}=\begin{pmatrix}\dfrac{1}{(1-r^2)\sigma_1^2} & \dfrac{-r}{(1-r^2)\sigma_1\sigma_2}\\ \dfrac{-r}{(1-r^2)\sigma_1\sigma_2} & \dfrac{1}{(1-r^2)\sigma_2^2}\end{pmatrix}.$$

于是， $n=2$ 元正态分布的密度函数可写为矩阵形式

$$p(\vec{x})=\frac{1}{(2\pi)^{\frac{2}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{a})^{\tau}B^{-1}(\vec{x}-\vec{a})\right\}. \tag{5.1}$$

注意，我们用 A^{τ} 表示矩阵 A 的转置阵. 密度函数的这种矩阵表示式 (5.1) 很容易推广到一般的 n 维情形.

定理 5.1 设

$$\vec{a}=\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix},\qquad B=\begin{pmatrix}b_{11}&\cdots&b_{1n}\\ \vdots& &\vdots\\ b_{n1}&\cdots&b_{nn}\end{pmatrix},\tag{5.2}$$

其中 B 为 $n\times n$ 正定对称阵. 则由 (5.1) 式给出的 $p(\vec{x})$ 满足

$$p(\vec{x})>0 \qquad 且 \qquad \int_{\Re^n}p(\vec{x})d\vec{x}=1.$$

以此 $p(\vec{x})$ 为密度函数的连续型分布称为 **n元正态分布**, 记作 $N(\vec{a},B)$.

证明: $p(\vec{x})>0$ 显然成立. 写对称阵 $B=LL^{\tau}$, 其中 L 仍非退化, $|L|=|B|^{1/2}>0$. 运用变数替换

$$\vec{y}=L^{-1}(\vec{x}-\vec{a})\tag{5.3}$$

可算得积分

$$\int_{\Re^n}p(\vec{x})\,d\vec{x}=\int_{\Re^n}(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\vec{y}^{\,\tau}\vec{y}\right\}d\vec{y}$$

$$=\left[\int_{\Re^1}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{y_k^2}{2}}\,dy_k\right]^n=1.$$

定理证毕.

如下定理给出 $N(\vec{a},B)$ 分布的特征函数.

定理 5.2 n 元正态分布 $N(\vec{a},B)$ 的特征函数为

$$f(\vec{t})=\exp\left\{i\vec{a}^{\,\tau}\vec{t}-\frac{1}{2}\vec{t}^{\,\tau}B\vec{t}\right\},\qquad \vec{t}\in\Re^n.\tag{5.4}$$

证明: 直接用特征函数的定义可算出 (5.1) 式中密度函数对应的特征函数

$$f(\vec{t})=\int_{\Re^n}(2\pi)^{-\frac{n}{2}}|B|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{a})^{\tau}B^{-1}(\vec{x}-\vec{a})+i\vec{t}^{\,\tau}\vec{x}\right\}d\vec{x}.$$

仍用变数替换 (5.3), 再配方可得

$$f(\vec{t}) = \int_{\Re^n} (2\pi)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2}\vec{y}^\tau \vec{y} + i\vec{t}^\tau (L\vec{y} + \vec{a}) \right\} d\vec{y}.$$

$$= \exp\left\{ i\vec{a}^\tau \vec{t} - \frac{1}{2}\vec{t}^\tau B\vec{t} \right\} \int_{\Re^n} (2\pi)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2}(\vec{y} - iL^\tau \vec{t})^\tau (\vec{y} - iL^\tau \vec{t}) \right\} d\vec{y}$$

最后令 $\vec{z} = \vec{y} - iL^\tau \vec{t}$ 便得

$$f(\vec{t}) = \exp\left\{ i\vec{a}^\tau \vec{t} - \frac{1}{2}\vec{t}^\tau B\vec{t} \right\} \left[\int_{\Re^1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_k^2}{2}} dz_k \right]^n$$

$$= \exp\left\{ i\vec{a}^\tau \vec{t} - \frac{1}{2}\vec{t}^\tau B\vec{t} \right\}.$$

(5.4) 式得证.

注意, 特征函数表达式 (5.4) 中已不含参数矩阵 B 求逆运算. 那么, 能不能将 n 元正态 $N(\vec{a}, B)$ 定义中 “B 为正定阵” 这一条件减弱为 “B 为半正定” 呢? 下面通过特征函数来实现这一扩充.

首先, 当 B 为对称正定阵时, 上述定理 5.1 与定理 5.2 保证了 (5.4) 式给出的 $f(\vec{t})$ 是一概率分布的特征函数. 如果减弱为 B 为半正定, 那么 (5.4) 式的 $f(\vec{t})$ 是否还是特征函数呢? 回答是肯定的. 事实上, 对半正定对称阵 B 及自然数 k, 令

$$B_k = B + \frac{1}{k}I,$$

其中 I 为 $n \times n$ 单位阵, 则每个 B_k 为对称正定阵. 通过 (5.4) 可定义 $N(\vec{a}, B_k)$ 的特征函数 $f_k(\vec{t})$. 注意到 $k \to \infty$ 时 $B_k \to B$, 而由指数函数的连续性知 $f_k(\vec{t}) \to f(\vec{t}) = \exp\{i\vec{a}^\tau \vec{t} - \frac{1}{2}\vec{t}^\tau B\vec{t}\}$, 且 $f(\vec{t})$ 连续. 则由连续性定理 (参见 4.1.3 节) 可知此 $f(\vec{t})$ 仍为 n 元特征函数.

定义 5.1 对于由 (5.2) 式给出的 n 维向量 $\vec{a}$ 及 $n \times n$ 对称半正定矩阵 B, 以 (5.4) 式中 $f(\vec{t})$ 为特征函数的概率分布称作 n 元正态分布, 记作 $N(\vec{a}, B)$.

由定理 5.2 知，当 B 正定 (即 $|B|>0$) 时，$N(\vec{a},B)$ 是 $\Re^n$ 上的一个连续型分布，其密度函数由 (5.1) 式给出. 当 $|B|=0$ 时，它不再是连续型的，而是退化到一个 r 维子空间上，其中 $r\,(<n)$ 为矩阵 B 的秩. 例如，当 $n=1$ 时，如果 $B=\sigma^2=0$, $N(a,0)$ 分布为单点 a 上的退化分布. 而对于 $n=2$ 维正态分布，当 $|B|=0$ 时则可以退化到 $\Re^2$ 中某条曲线或某个点上. 这样一来，由定义 5.1 给出的 n 元正态分布包括了它的退化情形，从而使正态分布族对某种极限运算封闭. 这将为今后讨论极限定理带来方便.

3.5.2 正态分布的性质

以下总设随机向量 $\vec{\xi}$ 服从 n 元正态分布，其特征函数如 (5.4) 式给出.

定理 5.3 边缘分布 设 $\vec{\xi_1}=(\xi_{s_1},\cdots,\xi_{s_k})^\tau$ 为 $\vec{\xi}$ 的任一子向量，则 $\vec{\xi_1}$ 服从 $N(\vec{a}_1,B_1)$ 分布，其中 $\vec{a}_1,B_1$ 分别为 $\vec{a},B$ 保留指标 $s_1,\cdots,s_k$ 对应的行和列所得的子矩阵. 特别，每个分量 ξ_k 服从 $N(a_k,b_{kk})$ 分布.

证明: 在 (5.4) 式中取 $t_j=0,\ j\neq s_1,\cdots,s_k$, 则得到 $\vec{\xi_1}$ 的边缘特征函数

$$E(\,e^{i\vec{t_1}^{\,\tau}\vec{\xi_1}})=\exp\{i\vec{a}_1^{\,\tau}\vec{t_1}-\frac{1}{2}\vec{t_1}^{\,\tau}B_1\vec{t_1}\},$$

即 $\vec{\xi_1}$ 服从 $N(\vec{a}_1,B_1)$ 分布. 定理得证.

定理 5.4 数字特征 $\vec{a}$ 为 $\vec{\xi}$ 的期望向量，而 B 为 $\vec{\xi}$ 的协方差阵. 即有

$$a_k=E(\xi_k),\qquad b_{jk}=\operatorname{cov}(\xi_j,\xi_k),\qquad j,k=1,2,\cdots,n. \tag{5.5}$$

证明: 由定理 5.3 的最后一论断及一维正态的性质可知，$a_k=E(\xi_k)$, 且 $b_{kk}=D(\xi_k)$. 只需再对 $j\neq k$ 证明 $\operatorname{cov}(\xi_j,\xi_k)=b_{jk}$. 事实上，由每个 $E|\xi_j|^2<+\infty$ 知 $E|\xi_j\xi_k|<+\infty$. 故用上节 (4.30) 式可得

$$E(\xi_j\xi_k)=\frac{1}{i^2}\frac{\partial^2 f(0,\cdots,0)}{\partial t_j\partial t_k}=b_{jk}+a_ja_k.$$

至此可得 $j \neq k$ 时

$$\mathrm{cov}(\xi_j, \xi_k) = E(\xi_j \xi_k) - E(\xi_j)E(\xi_k) = b_{jk}.$$

定理得证.

此定理表明，n 元正态分布由它的一与二阶矩完全决定.

定理 5.5 n 维正态随机向量的各分量相互独立，当且仅当它们两两不相关.

证明: 如果 $\vec{\xi}$ 的分量 $\xi_1, \cdots, \xi_n$ 相互独立，自然是两两不相关的. 反过来说，设每个 $j \neq k, b_{jk} = \mathrm{cov}(\xi_j, \xi_k) = 0$, 即 B 为对角阵. 那么，(5.4) 式的联合特征函数化为

$$f(t_1, \cdots, t_n) = \prod_{k=1}^{n} e^{ia_k t_k - \frac{1}{2} b_{kk} t_k^2} = \prod_{k=1}^{n} f_k(t_k).$$

由定理 5.3 知，其中 f_k 为 ξ_k 的边缘特征函数. 于是有 $\xi_1, \cdots, \xi_n$ 相互独立. 定理证完.

定理 5.6 设

$$\vec{\xi} = \begin{pmatrix} \vec{\xi}_1 \\ \vec{\xi}_2 \end{pmatrix},$$

其中 $\vec{\xi}_1, \vec{\xi}_2$ 为 $\vec{\xi}$ 的任一分块，并将 $\vec{a}$ 与 B 也相应地分块为

$$\vec{a} = \begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \tag{5.6}$$

则 $\vec{\xi}_1$ 与 $\vec{\xi}_2$ 独立的充分必要条件为 $B_{12} = 0$ (此时有 $B_{21} = 0$).

证明: 必要性 设 $\vec{\xi}_1$ 与 $\vec{\xi}_2$ 独立，则前者的任一分量 ξ_j 与后者的任一分量 ξ_k 独立. 从而 $b_{jk} = b_{kj} = 0$. 至此得 $B_{12} = 0$ 且 $B_{21} = 0$.

充分性 由分块矩阵的乘法知，当 $B_{12} = 0$ 且 $B_{21} = 0$ 时，$\vec{\xi} = (\vec{\xi}_1, \vec{\xi}_2)^\tau$ 的联合特征函数为

$$f(\vec{t}) = E\left[e^{i(\vec{t}_1{}^\tau \vec{\xi}_1 + \vec{t}_2^\tau \vec{\xi}_2)} \right]$$

$$= e^{i\vec{a}_1{}^\tau \vec{t}_1 - \frac{1}{2}\vec{t}_1{}^\tau B_{11}\vec{t}_1} \cdot e^{i\vec{a}_2{}^\tau \vec{t}_2 - \frac{1}{2}\vec{t}_2{}^\tau B_{22}\vec{t}_2},$$

右方恰为 $\vec{\xi}_1$ 与 $\vec{\xi}_2$ 边缘特征函数之积，故 $\vec{\xi}_1$ 与 $\vec{\xi}_2$ 独立．定理证完．

3.5.3 线性变换

仍设随机向量 $\vec{\xi}$ 服从 n 元正态分布 $N(\vec{a}, B)$. 考虑线性变换

$$\vec{\eta} = C\vec{\xi}, \tag{5.7}$$

其中 $C = [C_{jk}]$ 为一 $m \times n$ 矩阵．于是 $\vec{\eta}$ 为一 m 维随机向量．如下定理表明，正态分布随机向量的任何线性变换仍是正态分布随机向量，即正态性在线性变换下保持不变．

定理 5.7 (5.7) 式中 $\vec{\eta}$ 服从 $N(C\vec{a}, CBC^\tau)$ 分布．

证明: 仍以 $f(\vec{t})$ 表示 $\vec{\xi}$ 的特征函数，则 $\vec{\eta}$ 的特征函数为

$$g(\vec{s}) = E\left(e^{i\vec{s}^\tau \vec{\eta}}\right) = E\left(e^{i\vec{s}^\tau C\vec{\xi}}\right) = E\left(e^{i(C^\tau \vec{s})^\tau \vec{\xi}}\right)$$

$$= f(C^\tau \vec{s}) = e^{i(C\vec{a})^\tau \vec{s} - \frac{1}{2}\vec{s}^\tau (CBC^\tau)\vec{s}},$$

这正是 $N(C\vec{a}, CBC^\tau)$ 分布的特征函数．证毕．

推论 5.8 存在 n 维正交阵 C, 使 (5.7) 式中 $\vec{\eta}$ 各分量相互独立．

证明: 因为任何半正定对称阵 B, 总可通过正交变换 C, 使 CBC^τ 为对角型．由定理 5.5 知此时 $\vec{\eta}$ 各分量相互独立．

推论 5.9 若 $\vec{\xi}$ 各分量独立同方差，则对任何的 n 阶正交阵 C, (5.7) 式中 $\vec{\eta}$ 的各分量仍独立同方差．

证明: 此时 $\vec{\xi}$ 的协方差阵 $B = \sigma^2 I$ $(\sigma^2 \geq 0)$, 而对任何的 n 阶正交阵 C, $\vec{\eta}$ 的协方差阵 $CBC^\tau = C\sigma^2 IC^\tau = \sigma^2 I$. 即独立且同方差这一性质在正交变换下保持不变．

下面给出 n 元正态分布的一个重要等价条件．

定理 5.10 n 维随机向量 $\vec{\xi}$ 服从 $N(\vec{a},B)$ 分布，当且仅当对任何 n 维向量 $\vec{s}$，使得 $\eta=\vec{s}^{\tau}\vec{\xi}$ 有 1 维正态 $N(\vec{s}^{\tau}\vec{a},\vec{s}^{\tau}B\vec{s})$ 分布.

证明: 定理 5.7 中取 $C=\vec{s}^{\tau}$ 即得必要性. 下设对任意的 n 维向量 $\vec{s}$，随机变量 $\eta=\vec{s}^{\tau}\vec{\xi}$ 是 $N(\vec{s}^{\tau}\vec{a},\vec{s}^{\tau}B\vec{s})$ 分布，其特征函数

$$g(u)=E\left(e^{iu(\vec{s}^{\tau}\vec{\xi})}\right)=\exp\left\{iu(\vec{s}^{\tau}\vec{a})-\frac{1}{2}u^2\vec{s}^{\tau}B\vec{s}\right\}.$$

取 $u=1$ 便得

$$E\left(e^{i\vec{s}^{\tau}\vec{\xi}}\right)=\exp\left\{i\vec{s}^{\tau}\vec{a}-\frac{1}{2}\vec{s}^{\tau}B\vec{s}\right\}=f(\vec{s}),$$

这表明 $\vec{\xi}$ 的特征函数由 (5.4) 式给出，即服从 $N(\vec{a},B)$ 分布. 定理于是证完.

我们在第二章介绍 2 维概率分布时曾多次强调，随机向量的边缘分布不能决定联合分布. 在这里，同样要严格区分 "n 元正态随机向量" 与 "n 个一元正态随机变量" 这两个完全不同的概念. 按照定理 5.10, $\vec{\xi}$ 服从 n 元正态分布，等价于其分量的任何线性组合 $\vec{s}^{\tau}\vec{\xi}$ 均服从一维正态分布. 仅有各分量 ξ_k $(k=1,2,\cdots,n)$ 服从一维正态分布是远远不够的. 下面给出边缘分布是正态分布，而联合分布不是正态分布的一个例子.

例 5.1 设随机变量 ξ,η 独立，同服从标准正态分布. 令

$$\zeta=\begin{cases}|\eta|, & 当\ \xi>0,\\ -|\eta|, & 当\ \xi<0.\end{cases} \tag{5.8}$$

我们来证明 ζ 也是服从 $N(0,1)$ 分布的. 事实上，运用全概公式及 ξ 与 η 相互独立可知，当 $x>0$ 时 ζ 的分布函数

$$P\{\zeta<x\}=\frac{1}{2}P\{|\eta|<x\}+\frac{1}{2}P\{-|\eta|<x\}$$

$$=\frac{1}{2}\int_{-x}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt+\frac{1}{2}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x).$$

类似可得 $x<0$ 时 ζ 的分布函数也是 $\Phi(x)$. 至此得证 ζ 也是 $N(0,1)$ 分布的随机变量.

下面研究 (η,ζ) 的联合分布. 再次用全概公式及 ξ,η 独立可得

$$\begin{aligned}P\{\eta+\zeta=0\}&=\frac{1}{2}P\{\eta+|\eta|=0\}+\frac{1}{2}P\{\eta-|\eta|=0\}\\&=\frac{1}{2}P\{\eta<0\}+\frac{1}{2}P\{\eta>0\}=\frac{1}{2}.\end{aligned}$$

于是 $\eta+\zeta$ 不是 1 维正态分布的，故用定理 5.10 知 (η,ζ) 不服从 2 元正态分布. 可以证明 (ζ,ξ) 也不是 2 元正态分布随机向量. 留作练习.

但是由于已知 ξ 与 η 相互独立，其联合分布等于各自边缘分布的乘积，故 (ξ,η) 服从 2 元正态分布.

最后，我们介绍一个应用定理 5.7 的例题.

例 5.2 设 ξ_1,ξ_2,ξ_3 是相互独立的 $N(a,\sigma^2)$ 随机变量，而

$$\eta_1=\frac{1}{\sqrt{3}}(\xi_1+\xi_2+\xi_3),\qquad \eta_2=\frac{1}{\sqrt{2}}(\eta_1-\eta_2).$$

试求 $D(\eta_1\eta_2)$.

解: 可以把 η_1,η_2 的表达式代入，按定义计算 $\eta_1\eta_2$ 的方差. 但是可以用更简便的方法. 记

$$C=\begin{pmatrix}\frac{1}{\sqrt{3}}&\frac{1}{\sqrt{3}}&\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}&0\end{pmatrix},$$

则

$$\vec{\eta}=\begin{pmatrix}\eta_1\\\eta_2\end{pmatrix}=C\begin{pmatrix}\xi_1\\\xi_2\\\xi_3\end{pmatrix}.$$

用定理 5.7 知 $\vec{\eta}$ 服从 2 元正态分布，其均值向量与协方差矩阵分别为

$$C\begin{pmatrix}a\\a\end{pmatrix}=\begin{pmatrix}\sqrt{3}a\\0\end{pmatrix},\quad CBC^{\tau}=\begin{pmatrix}\sigma^2&0\\0&\sigma^2\end{pmatrix}.$$

因此 η_1 与 η_2 独立，η_1 服从 $N(\sqrt{3}a,\sigma^2)$ 分布，η_2 服从 $N(0,\sigma^2)$ 分布. 由此易得

$$D(\eta_1,\eta_2)=E(\eta_1^2\eta_2^2)-[E(\eta_1\eta_2)]^2$$

$$= E(\eta_1^2)E(\eta_2^2) - [E(\eta_1)E(\eta_2)]^2 = (\sigma^2 + 3a)\sigma^2.$$

3.5.4 习题

1. 设 (ξ, η) 服从 2 元正态分布，期望向量与协方差阵为

$$\vec{a} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}.$$

试求它的 (1) 联合密度函数； (2) 联合特征函数.

2. 若 $\vec{\xi} = (\xi_1, \cdots, \xi_n)^\tau$ 为 n 元正态分布的随机向量. 求证对任何常数 $b_0, b_1, \cdots, b_n$ 有

$$P\{b_1\xi_1 + \cdots + b_n\xi_n = b_0\} = 0 \text{ 或 } 1.$$

3. 设 (ξ_1, ξ_2) 是 2 元正态分布的随机向量，且对 $k, j = 1, 2$ 有 $E(\xi_k) = 0$ 与 $E(\xi_k\xi_j) = \sigma^2 > 0$. 求 $\eta = \frac{1}{\sigma^2}\xi_1\xi_2$ 的密度函数.

4. 试求例 5.1 中随机向量 (ζ, ξ) 的协方差阵，并由此证明它不服从 2 元正态分布

5. 设随机向量 (ξ, η) 服从 2 元正态分布，试导出 $\xi + \eta$ 与 $\xi - \eta$ 相互独立的充分必要条件.

6. 设 ξ 与 η 为相互独立的 $N(0, \sigma^2)$ 分布随机变量，试求 $a\xi + b\eta$ 与 $c\xi + d\eta$ 的相关系数. 这里 $\sigma^2 > 0$ 且 $ab \neq 0, cd \neq 0$.

7. 设随机变量 $\xi_1, \cdots, \xi_n$ 相互独立，都服从标准正态分布，试对 $m < n$ 求

$$\eta_1 = \sum_{k=1}^{n} \xi_k \quad \text{与} \quad \eta_2 = \sum_{k=1}^{m} \xi_k$$

的联合分布.

8. 设随机变量 $\xi_1, \cdots, \xi_n$ 相互独立，都服从 $N(0, \sigma^2)$ 分布. 取其平均值 $\overline{\xi} = \frac{1}{n}\sum_{k=1}^{n} \xi_k$.

(1) 试求 $\overline{\xi}$ 的分布；

(2) 求 $\overline{\xi}$ 与 ξ_1 的相关系数；

(3) 证明 $\overline{\xi}$ 与 $\eta = \sum_{k=1}^{n} (\xi_k - \overline{\xi})^2$ 独立.

第四章　极限定理

4.1　随机变量列的收敛性

4.1.1　作为可测函数列的收敛性

为了阐述概率论中的极限定理，必须先介绍随机变量序列的几种收敛性，扼要叙述有关概念、性质及它们之间的关系.

如前所述，按照概率论的公理系统，概率空间 $(\Omega,\mathcal{F},P)$ 上的随机变量就是样本空间 Ω 上关于 $\mathcal{F}$ 可测的实函数. 作为一般的实值可测函数的序列，随机变量列有相应的依测度 (概率) 收敛，几乎必然收敛和 r 阶平均收敛性. 先简述这三种收敛性的有关内容. 总设 ξ 与 $\xi_n(n\geq 1)$ 是给定概率空间 $(\Omega,\mathcal{F},P)$ 上的随机变量.

定义 1.1 如果有

$$\lim_{n} P\{|\xi_n(\omega)-\xi(\omega)|\geq\varepsilon\}=0,\qquad 任\quad \varepsilon>0, \tag{1.1}$$

则称 $\{\xi_n\}$ **依概率收敛** 到 ξ, 记作 $\xi_n\xrightarrow{P}\xi$.

直观地说，(1.1) 式表明，当 n 充分大时，ξ_n 与 ξ 有较大差异的概率可任意地小.

例 1.1 取区间 [0,1) 上的几何概率空间，即 $\Omega=[0.1)$, $\mathcal{F}$ 为区间 [0.1) 内的波雷尔集类，P 为勒贝格测度. 令

$$\eta_{11}(\omega)=1;$$

$$\eta_{21}(\omega)=\begin{cases}1, & \omega\in[0,1/2),\\ 0, & \omega\in[1/2,1);\end{cases}$$

$$\eta_{22}(\omega)=\begin{cases}1, & \omega\in[1/2,1),\\ 0, & \omega\in[0,1/2);\end{cases}$$

一般地，对每个自然数 k, 将 $[0,1)$ 等分为 k 个子区间，并取

$$\eta_{ki}(\omega)=\begin{cases}1, & \omega\in[(i-1)/k,i/k),\\ 0, & \text{其它}.\end{cases}\qquad i=1,2,\cdots,k.$$

定义

$$\xi_1=\eta_{11},\quad \xi_2=\eta_{21},\quad \xi_3=\eta_{22},\quad \xi_4=\eta_{31},\cdots.$$

易见 $\{\xi_n\}$ 为随机变量序列，由于任何 $\varepsilon>0$ 有

$$P\{|\eta_{ki}(\omega)|\geq\varepsilon\}=P\{[(i-1)/k,i/k)\}=1/k\longrightarrow 0.$$

故由定义 1.1 知 $\xi_n\xrightarrow{P}0$.

定义 1.2 如果有

$$P\{\omega:\lim_n\xi_n(\omega)=\xi(\omega)\}=1,\tag{1.2}$$

则称随机变量列 $\{\xi_n\}$**几乎必然收敛**到 ξ, 记作 $\xi_n\xrightarrow{a.s.}\xi$.

(1.2) 式表明，对于 Ω 上的几乎所有样本点 ω, 数列 $\{\xi_n(\omega)\}$ 收敛到 $\xi(\omega)$. 采用分析中数列收敛的语言，此时对任何 $\varepsilon>0$, 都有 $k\geq 1$, 使当 $n\geq k$ 时，永有 $|\xi_n(\omega)-\xi(\omega)|<\varepsilon$. 于是，随机变量序列 $\{\xi_n\}$ 的收敛点集可写成

$$E=\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left\{\omega:|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right\}\in\mathcal{F}.$$

而 $\xi_n\xrightarrow{a.s.}\xi$ 的等价条件是 $P(E)=1$ 或 $P(\overline{E})=0$. 进一步地有

定理 1.1 $\xi_n\xrightarrow{a.s.}\xi$ 的充分必要条件是

$$\lim_k P\{\bigcup_{n=k}^{\infty}[|\xi_n-\xi|\geq\varepsilon]\}=0,\qquad \text{任}\ \ \varepsilon>0.\tag{1.3}$$

证明: 对任何自然数 k,m 记

$$A_{k,m}=\bigcup_{n=k}^{\infty}\left\{|\xi_n-\xi|\geq\frac{1}{m}\right\}\in\mathcal{F}.$$

易见 (1.3) 式等价于

$$\lim_k P(A_{k,m}) = 0, \qquad 任 \quad m \geq 1. \tag{1.4}$$

注意到 $k \to \infty$ 时 $A_{k,m} \downarrow \bigcap_{k=1}^{\infty} A_{k,m}$. 由概率的上连续性知 (1.4) 等价于

$$P(\bigcap_{k=1}^{\infty} A_{k,m}) = 0, \qquad 任 \quad m \geq 1. \tag{1.5}$$

至此只要再证 (1.5) 式是 $\xi_n \xrightarrow{a.s.} \xi$ 的充分必要条件. 这一点由

$$\overline{E} = \bigcup_{m=1}^{\infty} \bigcap_{k=1}^{\infty} A_{k,m}$$

及

$$P(\bigcap_{k=1}^{\infty} A_{k,m}) \leq P(\overline{E}) \leq \sum_{m=1}^{\infty} P(\bigcap_{k=1}^{\infty} A_{k,m})$$

便可推出. 定理得证.

推论 1.2 若 $\xi_n \xrightarrow{a.s.} \xi$, 则 $\xi_n \xrightarrow{P} \xi$.

证明: 此推论由任 $\varepsilon > 0$

$$\{|\xi_n - \xi| \geq \varepsilon\} \subset \bigcup_{k=n}^{\infty} \{|\xi_k - \xi| \geq \varepsilon\}$$

及 (1.3) 式立即得证.

推论 1.2 说明, 几乎必然收敛性比依概率收敛性要强. 此推论的逆不真, 上述例 1.1 便是例证. 实际上, 此时对一切 ω 及无论多么大的 k, 总有 $n_1, n_2 > k$ 使 $\xi_{n_1} = 1$ 而 $\xi_{n_2} = 0$. 就是说, 例 1.1 中的随机变量列对一切 ω 不收敛. 即其收敛点集 $E = \emptyset$. 可以证明, 每个依概率收敛的随机变量列必存在某个子序列几乎必然收敛到同一随机变量.

定义 1.3 设 $r > 0$ 为常数, 如果随机变量 ξ 与 $\xi_n (n \geq 1)$

的 r 阶矩皆有限，并且有

$$\lim_n E|\xi_n - \xi|^r = 0, \tag{1.6}$$

则称 $\{\xi_n\}$ 为 **r 阶平均收敛** 到 ξ，简称 r 阶收敛，记作 $\xi_n \xrightarrow{L_r} \xi$. 当 $r = 1$ 时可称作**平均收敛**，$r = 2$ 时称为**均方收敛**.

为给出 r 阶平均收敛与依概率收敛之间的关系，我们需要如下有用的不等式.

引理 1.3 马尔科夫不等式 设随机变量 ξ 的 r 阶矩有限，则有

$$P\{|\xi| \geq \varepsilon\} \leq \frac{1}{\varepsilon^r} E|\xi|^r, \qquad 任 \ \ \varepsilon > 0. \tag{1.7}$$

证明: 设 ξ 的分布函数为 $F(x)$，则对任何 $\varepsilon > 0$ 有

$$E|\xi|^r = \int_{\Re} |x|^r dF(x) \geq \int_{(|x| \geq \varepsilon)} |x|^r dF(x)$$

$$\geq \varepsilon^r \int_{(|x| \geq \varepsilon)} dF(x) = \varepsilon^r P\{|\xi| \geq \varepsilon\},$$

这就证明了 (1.7) 式.

定理 1.4 若 $\xi_n \xrightarrow{L_r} \xi$，则 $\xi_n \xrightarrow{P} \xi$.

证明: 设 $\xi_n \xrightarrow{L_r} \xi$，对 $\xi_n - \xi$ 用马尔科夫不等式 (1.7) 便得，

$$P\{|\xi_n - \xi| \geq \varepsilon\} \leq \frac{1}{\varepsilon^r} E|\xi_n - \xi|^r \longrightarrow 0, \qquad 任 \ \ \varepsilon > 0.$$

这表明 $\xi_n \xrightarrow{P} \xi$. 证毕.

此定理的逆亦不真，即 r 阶平均收敛确实比依概率收敛要强. 且看如下的例.

例 1.2 仍取区间 [0,1) 上的几何概率空间 $(\Omega, \mathcal{F}, P)$. 固定 $r > 0$，令

$$\xi_n(\omega) = \begin{cases} n^{1/r}, & \omega \in [0, 1/n), \\ 0, & \omega \in [1/n, 1), \end{cases} \qquad \xi(\omega) \equiv 0.$$

易见对一切 ω 有 $\xi_n(\omega) \to \xi(\omega)$. 用定理 1.1 知 $\xi_n \xrightarrow{P} \xi$. 但是 $E|\xi_n - \xi|^r = E|\xi_n|^r = 1 \not\to 0$, 即 $\{\xi_n\}$ 不是 r 阶平均收敛到 ξ 的.

由上可知, 几乎必然收敛与 r 阶平均收敛都比依概率收敛确实地强. 但是它们二者却没有蕴含关系, 上述例 1.1 与例 1.2 恰好证明这一点.

4.1.2 弱收敛

本节将借助分布函数引进随机变量序列的另一种收敛性. 它比依概率收敛性更弱, 故称作弱收敛. 同时, 由于它等价于相应的特征函数序列的逐点收敛, 因而在概率论的极限定理中居重要的地位.

定义 1.4 设 $F(x), F_n(x)(n \geq 1)$ 均为实函数. 如果有

$$\lim_n F_n(x) = F(x), \qquad 任 \ x \in C_F, \tag{1.8}$$

其中 C_F 为函数 F 的连续点集,则称 $\{F_n\}$ **弱收敛**到 F, 记作 $F_n \xrightarrow{w} F$.

这样定义的收敛性确实太弱. 例如, 任意函数列 $\{F_n\}$ 均弱收敛于一个点点不连续的函数 F. 又如, 当 $F_n \xrightarrow{w} F$ 时, 可在 C_F 之外任意改动 F 的值而保持弱收敛性, 就是说, 弱收敛的极限函数不唯一. 但是, 如果再设 $F(x)$ 为连续函数, 或者象分布函数那样单调左连续, 则弱收敛的含义将大大增强.

定义 1.5 设随机变量 ξ_n 与 ξ 分别有分布函数 $F_n(x)$ 与 $F(x)$, 且 $F_n \xrightarrow{w} F$, 则称随机变量列 $\{\xi_n\}$**依分布收敛**到 ξ, 仍记作 $\xi_n \xrightarrow{w} \xi$.

例 1.3 对一切自然数 n 令

$$F_n(x) = \begin{cases} 0, & x \leq -1/n, \\ 1, & x > -1/n, \end{cases} \qquad F(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x > 0. \end{cases}$$

易见 $F_n \xrightarrow{w} F$. 注意到 F_n 与 F 分别是随机变量 $\xi_n = -1/n$ 与 $\xi = 0$ 的分布函数, 故按定义知 $\{\xi_n\}$ 依分布收敛到 ξ.

如果在定义中，不允许在极限函数的间断点上有例外，那么连本例中处处收敛的常数列 $\{\xi_n\}$ 也不具有这种收敛性. 可见，弱收敛定义中排除极限函数的间断点是必要的. 另一方面，在 F_n 与 F 都是分布函数的情形，$\overline{C}_F$ 至多可列，由左连续性知，$\overline{C}_F$ 上的 F 由 C_F 上的值完全确定. 就是说，此时弱收敛的极限函数是唯一的. 因此，这种允许在 F 的间断点上有例外的作法又是可行的. 下面证明这种收敛性还有更重要的性质.

定理 1.5 若 $\xi_n \xrightarrow{P} \xi$, 则 $\xi_n \xrightarrow{\mathrm{w}} \xi$.

证明: 设 $F_n(x)$ 与 $F(x)$ 为 ξ_n 与 ξ 相应的分布函数，对任意的 $y < x$, 由

$$\{\xi < y\} = \{\xi < y, \xi_n < x\} \bigcup \{\xi < y, \xi_n \geq x\}$$

$$\subset \{\xi_n < x\} \bigcup \{|\xi_n - \xi| \geq x - y\}$$

推知

$$F(y) \leq F_n(x) + P\{|\xi_n - \xi| \geq x - y\}.$$

再令 $n \to \infty$, 用 $\xi_n \xrightarrow{P} \xi$ 可得

$$F(y) \leq \varliminf_n F_n(x).$$

类似对任意 $x < z$ 有

$$\varlimsup_n F_n(x) \leq F(z).$$

现设 $x \in C_F$, 取 $y \uparrow x$ 且 $z \downarrow x$ 便得到

$$F(x) \leq \varliminf_n F_n(x) \leq \varlimsup_n F_n(x) \leq F(x).$$

于是

$$F(x) = \lim_n F_n(x), \qquad 任 \ \ x \in C_F.$$

此即 $\{\xi_n\}$ 依分布收敛于 ξ. 定理得证.

本定理的逆不成立.

例 1.4 设 ξ,η 独立，同服从 $p=1/2$ 的贝努里分布，对一切 $n\geq 1$ 取 $\xi_n=\eta$, 则由 ξ_n 与 ξ 同分布知 $\{\xi_n\}$ 依分布收敛到 ξ. 但是有

$$P\{|\xi_n-\xi|=1\}=P\{|\eta-\xi|=1\}=\frac{1}{2}.$$

可见 $\{\xi_n\}$ 不是依概率收敛到 ξ 的.

但是当极限函数为常数 c 时，依分布收敛与依概率收敛等价.

定理 1.6 若 $\xi_n \xrightarrow{W} c$, 则 $\xi_n \xrightarrow{P} c$.

证明: 因为常数 c 的分布函数只在 c 处不连续，故对任何 $\varepsilon>0$ 有

$$\begin{aligned}P\{|\xi_n-c|\geq\varepsilon\}&=P\{\xi_n\geq c+\varepsilon\}+P\{\xi_n\leq c-\varepsilon\}\\&=1-F_n(c+\varepsilon)+F_n(c-\varepsilon+0)\longrightarrow 0,\end{aligned}$$

这表明 $\xi_n \xrightarrow{P} c$ 成立. 证毕.

弱收敛还有如下形式上较弱的等价形式.

定理 1.7 设 $F(x),F_n(x)(n\geq 1)$ 均为分布函数，则 $F_n \xrightarrow{W} F$ 的充分必要条件是，对于 C_F 的某个稠子集 D 有

$$\lim_n F_n(x)=F(x),\qquad \text{任}\ \ x\in D. \tag{1.9}$$

证明: 由 $D\subset C_F$ 立得必要性. 下设 (1.9) 式成立. 对任何 $x\in C_F$, 取 $y<x<z$ 且 $y,z\in D$, 则有

$$F_n(y)\leq F_n(x)\leq F_n(z).$$

上式令 $n\to\infty$, 用 (1.9) 式得

$$F(y)=\lim_n F_n(y)\leq \varliminf_n F_n(x)\leq \varlimsup_n F_n(x)\leq \lim_n F_n(z)=F(z).$$

再令 $y\uparrow x$ 及 $z\downarrow x$ 便得证 (1.8) 式. 即 $F_n \xrightarrow{W} F$. 证毕.

4.1.3 连续性定理

本节的主要内容是证明分布函数列的弱收敛等价于相应特征函数列的逐点收敛. 这意味着由特征函数定义及其反演公式所建立起的分布函数与特征函数间的一一对应有某种连续性, 故称作连续性定理. 叙述中涉及一些著名的分析结果, 它们的内容十分丰富, 证明也较复杂. 有些读者可以着重在理解这些定理的含义及应用上, 而忽略某些证明细节.

定理 1.8 海莱第一定理 任何 $\Re$ 上的分布函数列都存在弱收敛于某个有界, 非降且左连续函数的子序列.

证明: 以 $D=\{r_1,r_2,\cdots\}$ 表 $\Re$ 中的有理数集. 对于 r_1, 有界数列 $\{F_n(r_1)\}$ 必存在收敛的子列 $\{F_{1,n}(r_1)\}$, 其极限记为 $G(r_1)$. 对于 r_2, 有界数列 $\{F_{1,n}(r_2)\}$ 也必存在收敛的子列 $\{F_{2,n}(r_2)\}$, 其极限记为 $G(r_2)$. 由于 $\{F_{2,n}\}$ 是 $\{F_{1,n}\}$ 的子列, 故同时有

$$\lim_n F_{2,n}(r_1)=G(r_1)\quad 与 \quad \lim_n F_{2,n}(r_2)=G(r_2).$$

循此下去, 可得序列 $\{F_{m,n}(x)\}$, 使每个 m, $\{F_{m,n}\}$ 是 $\{F_{m-1,n}\}$ 的子序列, 从而对一切 $m\ge 1$ 有

$$\lim_n F_{m,n}(r_k)=G(r_k),\qquad k=1,\cdots,m. \tag{1.10}$$

至此我们得到 $\{F_n\}$ 的如下子序列

$$\begin{array}{l}
F_{1,1},\ F_{1,2},\ \cdots,\ F_{1,n},\ \cdots\\
F_{2,1},\ F_{2,2},\ \cdots,\ F_{2,n},\ \cdots\\
\cdots\quad \cdots\quad \cdots\\
F_{m,1},\ F_{m,2},\ \cdots,\ F_{m,n},\ \cdots\\
\cdots\quad \cdots\quad \cdots
\end{array}$$

使得 (1.10) 式成立. 选取这个阵列的对角线构成的 $\{F_n\}$ 的子列 $\{F_{n,n}\}$. 由于此 $\{F_{n,n}\}$ 是 $\{F_{1,n}\}$ 的子列, 故有 $\lim\limits_n F_{n,n}(r_1)=G(r_1)$. 类似地, 对 $n\geq 2$, 由于 $\{F_{n,n}\}$ 又是 $\{F_{2,n}\}$ 的子列, 故有 $\lim\limits_n F_{n,n}(r_2)=G(r_2)$. 一般地, 对 $n\geq k$, 由于 $\{F_{n,n}\}$ 是 $\{F_{k,n}\}$ 的子列, 故有 $\lim\limits_n F_{n,n}(r_k)=G(r_k)$. 总之 $\{F_n\}$ 的这个子列 $\{F_{n,n}\}$ 有性质:

$$\lim_n F_{n,n}(r_k)=G(r_k), \qquad 任\ \ r_k\in D. \tag{1.11}$$

且 D 上的极限函数 G 也是非降及有界的.

将 D 上的 G 扩充到 $\Re$ 上, 取

$$F(x)=\sup_{r_k<x} G(r_k). \tag{1.12}$$

则 F 仍是非降且有界的. 且由于在 D 上 F 与 G 相等, 故由 (1.11) 式及定理 1.7 知 $F_{n,n}\xrightarrow{\text{w}}F$. 余下只需再证明 F 左连续. 事实上由 (1.12) 知, 对任何 $\varepsilon>0$, 总存在 $r_k<x$ 使 $G(r_k)>F(x)-\varepsilon$, 且对于一切 $y\in(r_k,x)$ 有

$$F(y)\geq G(r_k)>F(x)-\varepsilon.$$

上式先令 $y\uparrow x$, 再令 $\varepsilon\downarrow 0$ 便得 $F(x-0)\geq F(x)$. 此表明 (1.12) 式所定义的函数 F 是左连续的. 定理证毕.

定理证明中所用选取子序列的方法称作**对角线法**.

不能断定海莱第一定理中的极限函数 F 为分布函数. 事实上, 取 $\xi_n=n\ (n\geq 1)$, 则对应的分布函数 $F_n\xrightarrow{\text{w}}0$, 极限函数不是分布函数.

定理 1.9 海莱第二定理 设分布函数列 $\{F_n\}$ 弱收敛于分布函数 $F(x)$, 则对任何有界连续函数 φ 有

$$\int_{\Re}\varphi(x)\,dF_n(x)\longrightarrow\int_{\Re}\varphi(x)\,dF(x). \tag{1.13}$$

证明: 按 2.7.5 节的随机变量存在性定理 (定理 7.9), 取 $\Omega=(0,1)$, $\mathcal{F}$ 为区间 (0,1) 中的波雷尔集类, P 为勒贝格测度. 再在概率空间 $(\Omega,\mathcal{F},P)$ 上取

$$\xi_n=F_n^{-1}(\omega),\quad n\geq 1\quad \text{及}\quad \xi=F^{-1}(\omega),\quad \omega\in\Omega$$

则随机变量 ξ_n 与 ξ 分别以 $F_n(x)$ 及 $F(x)$ 为其分布函数. 注意, 分布函数 $\omega=F(x),x\in\Re$ 与其单调逆 $x=F^{-1}(\omega),\omega\in(0,1)$ 都是单调非降的, 且有相同的图形 (参见 2.7.5 节图 7.1). 故由 $F_n(x)\to F(x),x\in C_F$ 可推出 $F_n^{-1}(\omega)\to F^{-1}(\omega),\omega\in C_{F^{-1}}$. 也就是说, 所构造的随机变量列 $\{\xi_n\}$ 几乎必然收敛到 ξ. 在 $C_{F^{-1}}(\omega)$ 外改取 ξ_n 与 ξ 均为 0(这不影响它们的分布), 则 $\{\xi_n\}$ 处处收敛到 ξ. 从而对有界连续函数 φ, 序列 $\{\varphi(\xi_n)\}$ 收敛到 $\varphi(\xi)$. 用勒贝格控制收敛定理可得

$$E[\varphi(\xi_n)]\longrightarrow E[\varphi(\xi)].$$

此即待证的 (1.13) 式. 证完.

至此我们已准备好叙述本节的主要结果.

定理 1.10 连续性定理 分布函数列 $\{F_n(x)\}$ 弱收敛到分布函数 $F(x)$ 的充分必要条件是, 相应的特征函数列 $\{f_n(t)\}$ 逐点收敛到相应的特征函数 $f(t)$.

证明: 必要性 设 $F_n\xrightarrow{W}F$, 对有界连续函数 $\cos tx$ 与 $\sin tx$ 分别用定理 1.9 便得, 当 $n\to\infty$ 时对一切 $t\in\Re$ 有

$$\begin{aligned}f_n(t)&=\int_{\Re}\cos tx\,dF_n(x)+i\int_{\Re}\sin tx\,dF_n(x)\\&\longrightarrow\int_{\Re}\cos tx\,dF(x)+i\int_{\Re}\sin tx\,dF(x)=f(t).\end{aligned}$$

充分性 据定理 1.8 知, 分布函数列 $\{F_n\}$ 必存在子序列 $\{F_{n_k}\}$, 使当 $k\to\infty$ 时 $F_{n_k}\xrightarrow{W}\widehat{F}$. 其中极限函数 $\widehat{F}$ 是 $\Re$ 上非降左连续函数, 且有界: $\widehat{F}(-\infty)\geq 0$, $\widehat{F}(+\infty)\leq 1$. 下证此二式均取等号, 即 $\widehat{F}$ 为分布函数.

如若不然，有

$$a = \widehat{F}(+\infty) - \widehat{F}(-\infty) < 1. \tag{1.14}$$

那么，一方面由 $f(0) = 1$ 及 $f(t)$ 连续知，对满足 $0 < \varepsilon < 1 - a$ 的任意 ε, 存在充分小的正数 τ, 使

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)dt\right| > 1 - \frac{\varepsilon}{2} > a + \frac{\varepsilon}{2}. \tag{1.15}$$

另一方面，既然 $F_{nk} \xrightarrow{w} \widehat{F}$, 由 (1.14) 式知可选取 $b > 4/\varepsilon\tau$, 使 $-b$ 与 b 皆为 $\widehat{F}$ 的连续点，且存在自然数 K, 使当 $k \geq K$ 时有

$$a_k = F_{n_k}(b) - F_{n_k}(-b) < a + \frac{\varepsilon}{4}.$$

再由

$$\left|\int_{-\tau}^{\tau} e^{itx}dt\right| \leq 2\tau$$

及 $|x| > b$ 时有

$$\left|\int_{-\tau}^{\tau} e^{itx}dt\right| = \left|\frac{2}{x}\sin tx\right| \leq \frac{2}{b}$$

便可得到

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f_{n_k}(t)dt\right| = \frac{1}{2\tau}\left|\int_{-\infty}^{+\infty}\left[\int_{-\tau}^{\tau} e^{itx}dt\right]dF_{n_k}(x)\right|$$

$$\leq \frac{1}{2\tau}\left|\int_{(-b,b)}\left[\int_{-\tau}^{\tau} e^{itx}dt\right]dF_{n_k}(x)\right| + \frac{1}{2\tau}\left|\int_{(|x|\geq b)}\left[\int_{-\tau}^{\tau} e^{itx}dt\right]dF_{n_k}(x)\right|$$

$$\leq a_k + \frac{1}{b\tau} \leq a_k + \frac{\varepsilon}{4} < a + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = a + \frac{\varepsilon}{2},$$

这与 (1.15) 式矛盾. 至此得证 $\{F_n\}$ 的子列 $\{F_{n_k}\}$ 弱收敛到分布函数 $\widehat{F}$. 对此运用已得证的必要性知 $\widehat{F}$ 所对应的特征函数为 f. 再用唯一性定理可推出 $\widehat{F} = F$.

最后证明分布函数列 $\{F_n\}$ 也弱收敛到 F. 仍然用反证法. 如若不然，必存在 F 的连续点 x_0, 使 $F_n(x_0) \not\to F(x_0)$. 于是有界数列

$\{F_n(x_0)\}$ 必含收敛子列 $\{F_{m_k}(x_0)\}$. 其极限值 $F_{m_k}(x_0) \to F^*(x_0) \neq F(x_0)$. 对分布函数序列 $\{F_{m_k}\}$ 运用定理 1.8, 又存在子列 $\{F_{m_{k_l}}\}$ 使 $F_{m_{k_l}} \xrightarrow{W} F^*$. F^* 与前述 F 至少在 x_0 上不同. 但是重复上述论证可知 F^* 也应当是与 f 对应的分布函数, 由唯一性定理知 $F^* = F$, 这导出矛盾. 定理证完.

我们指出, 这里只叙述了这些著名定理的最必需的部分. 它们的实际内容要丰富得多. 比如, 定理 1.9 的逆也成立. 其它定理也可以明显地加强, 有的条件可以减弱, 有的结论可以增强. 限于篇幅, 不一一详述.

可将随机变量序列之几种收敛性间的蕴含关系总结如下:

几乎必然收敛	$\Rightarrow$	依概率收敛	$\Rightarrow$	分布函数的弱收敛
		$\Uparrow$		$\Updownarrow$
		r阶平均收敛		特征函数逐点收敛

4.1.4 习题

1. 求证例 1.1 中的随机变量序列是 r 阶平均收敛的.

2. 设 $\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 求证 (1) $\xi_n + \eta_n \xrightarrow{P} \xi + \eta$; (2) $\xi_n\eta_n \xrightarrow{P} \xi\eta$.

3. 设 $\xi_n \xrightarrow{P} a\,(a \neq 0)$, 求证 $\xi_n^{-1} \xrightarrow{P} a^{-1}$.

4. 试证 $\xi_n \xrightarrow{P} \xi$ 的充分必要条件是 $E[\frac{|\xi_n - \xi|}{1+|\xi_n - \xi|}] \to 0$.

5. 试证依概率收敛的随机变量序列必存在几乎必然收敛到同一极限的子序列.

6. 设 $\{\xi_n\}$ 为非负随机变量的单调下降序列, 且 $\xi_n \xrightarrow{P} 0$, 求证 $\xi_n \xrightarrow{a.s.} 0$.

7. 设对于一切自然数 $n \geq 1$, ξ_n 服从柯西分布, 其密度函数为

$$p_n(x) = \frac{n}{\pi(1 + n^2x^2)},$$

求证 $\xi_n \xrightarrow{P} 0$.

8. 设 $\{\xi_n\}$ 为独立随机变量序列, 它们有共同的密度函数

$$p(x) = e^{-(x-a)}, \quad x > a,$$

令 $\eta_n = \max\{\xi_1, \cdots, \xi_n\}$, 求证 $\eta_n \xrightarrow{P} a$.

9. 设正态分布函数列 $\{F_n(x)\}$ 弱收敛到分布函数 $F(x)$, 求证 $F(x)$ 也是正态分布函数, 且相应的期望与方差也收敛.

10. 设 $\{\xi_k\}$ 为独立随机变量序列, 它们都服从 -1 与 1 两点上的等可能分布. 试证 $\eta_n = \sum\limits_{k=1}^{n} \xi_k/2^k$ 依分布收敛到区间 $(-1,1)$ 上的均匀分布随机变量.

11. 设 $\{\xi_n\}$ 依分布收敛到 ξ, $\{\eta_n\}$ 依分布收敛到常数 a, 求证 $\{\xi_n + \eta_n\}$ 依分布收敛到 $\xi + a$.

12. 设 $\{\xi_n\}$ 依分布收敛到 ξ, $\{\eta_n\}$ 依分布收敛到 0. 求证 $\{\xi_n\eta_n\}$ 依分布收敛到 0.

13. 设 $\{\xi_n\}$ 依分布收敛到 ξ, 数列 $a_n \to a$, $b_n \to b$. 求证 $\{a_n\xi_n + b_n\}$ 依分布收敛到 $a\xi + b$.

14. 假定 $\{\xi_n\}$ 依概率收敛到 ξ, 求证对任意连续函数 g, 序列 $\{g(\xi_n)\}$ 依概率收敛到 $g(\xi)$. 再证将上述"依概率"均改为"依分布"命题仍真.

15. 设分布函数列 $\{F_n(x)\}$ 弱收敛到连续的分布函数 $F(x)$, 求证这种收敛对 $x \in \Re$ 是一致的.

4.2 大数定律

4.2.1 定义与马尔科夫条件

在本书的开头, 我们曾经用"频率的稳定性"引出概率这个基本概念. 许多试验结果表明, 虽然一次随机试验中某确定事件发生与否不能预言, 但是如果在相同条件下大量重复这个试验, 则此事件发生的频率会稳定在某个值附近. 这说明, 在一定条件下各事件出现的可能性的大小是客观存在的, 可以用上述频率的稳定值来度量, 这就是

事件的概率. 频率的稳定性呈现在大量重复试验中, 历史上把这个试验次数很大时出现的规律称作大数定律.

后来我们又引入了贝努里概型来刻画独立重复试验. 将一成功(即事件 A 发生) 概率为 p 的试验独立重复 n 次, 其中成功 μ_n 次, 则 μ_n 是二项分布随机变量. $E(\mu_n)=np$, $D(\mu_n)=npq$. 因此成功的频率 μ_n/n 也是随机变量, 其期望为 p 与 n 无关, 且方差 pq/n 当 $n\to\infty$ 时趋于 0. 熟知, 方差为 0 的随机变量恒等于它的期望, 所以 $n\to\infty$ 时频率 μ_n/n 应以概率 p 为极限. 另一方面, 可以写 $\mu_n=\sum\limits_{k=1}^{n}\xi_k$, 其中 $\xi_1,\cdots,\xi_n$ 相互独立, 具有相同的贝努里分布. 至此, 问题化为研究 $n\to\infty$ 时 ξ 的平均值序列 $\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 的极限行为. 鉴于已在上节讨论过随机变量列的各种收敛性, 因此我们可以给出大数定律的严格定义.

定义 2.1 设 $\{\xi_n\}$ 为随机变量序列, 它们都有有限的数学期望 $E(\xi_n)$. 如果

$$\frac{1}{n}\sum_{k=1}^{n}[\xi_k-E(\xi_k)]\xrightarrow{P}0, \tag{2.1}$$

则称 $\{\xi_n\}$**满足大数定律**. 如果进一步有

$$\frac{1}{n}\sum_{k=1}^{n}[\xi_k-E(\xi_k)]\xrightarrow{a.s.}0, \tag{2.2}$$

则称 $\{\xi_n\}$**满足强大数定律**.

由上节推论 1.2 知, (2.2) 式蕴含 (2.1) 式, 故大数定律有强弱之分. 习惯上总把弱大数定律的"弱"字省略. 以下将针对不同场合分别给出大数定律 (2.1) 与强大数定律 (2.2) 式成立的条件. 本节先介绍马尔科夫条件及其推论.

与普通数列的极限类似, 要证明随机变量序列收敛, 必须估计它与极限随机变量的差. 因而, 建立适当的不等式, 常常是证明收敛定理的前奏. 上节引理 1.3 已经给出了马尔科夫不等式, 它在 $r=2$ 时的特例便是著名的切贝谢夫不等式.

引理 2.1 切贝谢夫不等式 若随机变量 ξ 有有限的方差，则

$$P\{|\xi - E(\xi)| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} D(\xi), \qquad 任 \ \ \varepsilon > 0. \tag{2.3}$$

如下大数定律是本节的主要结果.

定理 2.2 设 $\{\xi_n\}$ 是方差有限的随机变量序列，如果有

$$\frac{1}{n^2} D(\sum_{k=1}^{n} \xi_k) \longrightarrow 0, \tag{2.4}$$

则 $\{\xi_n\}$ 满足大数定律.

证明: 由切贝谢夫不等式及 (2.4) 式立得，对任意 $\varepsilon > 0$ 有

$$P\{|\frac{1}{n}\sum_{k=1}^{n}[\xi_k - E(\xi_k)]| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2 n^2} D(\sum_{k=1}^{n} \xi_k) \longrightarrow 0,$$

即 (2.1) 式成立，定理得证.

(2.4) 式通常称为**马尔科夫条件**. 它是大数定律的一个容易验证的充分条件.

推论 2.3 切贝谢夫大数律 若序列 $\{\xi_n\}$ 两两不相关且方差有界: $D(\xi_n) \leq C$ $(n \geq 1)$, 则 $\{\xi_n\}$ 满足大数定律.

证明: 在所给的条件下，(2.4) 式的左方

$$\frac{1}{n^2} D(\sum_{k=1}^{n} \xi_k) = \frac{1}{n^2} \sum_{k=1}^{n} D(\xi_k) \leq \frac{C}{n} \longrightarrow 0,$$

即马尔科夫条件满足从而大数定律成立. 证完.

推论 2.4 贝努里大数律 设 μ_n 为 n 重贝努里试验中成功次数，则当 $n \to \infty$ 时有

$$\frac{\mu_n}{n} \xrightarrow{P} p. \tag{2.5}$$

证明: 如前所述，可写 $\mu_n = \sum_{k=1}^{n} \xi_k$，其中 $\{\xi_k\}$ 为独立的贝努里序列，且 $E(\xi_k) = p$, $D(\xi_k) = pq \le 1$. 因此推论 2.3 的条件满足，从而大数定律成立. 证毕.

此推论就是"频率以概率为其稳定值"的严格数学刻画.

4.2.2 若干引理

在进一步讨论大数定律成立的其它条件之前，本节先做一些必要的准备. 首先给出著名的波雷尔－康特里引理. 回忆事件列 $\{A_n\}$ 的上限集与下限集

$$\overline{\lim_{n}} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{有无限多个A_n发生\}, \tag{2.6}$$

$$\underline{\lim_{n}} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n = \{至多有限个A_n不发生\}. \tag{2.7}$$

引理 2.5 波雷尔－康特里引理

(1) 若 $\sum_{n=1}^{\infty} P(A_n) < +\infty$, 则 $P(\overline{\lim_{n}} A_n) = 0$;

(2) 设事件列 $\{A_n\}$ 相互独立，则 $\sum_{n=1}^{\infty} P(A_n) = +\infty$ 的充分必要条件是 $P(\overline{\lim_{n}} A_n) = 1$.

证明: (1) 记 $B_k = \bigcup_{n=k}^{\infty} A_n$, 由 (2.6) 式知 $B_k \downarrow \overline{\lim_{n}} A_n$. 据概率的上连续性，次可加性及 $\sum_{n=1}^{\infty} P(A_n) < +\infty$ 可得

$$P(\overline{\lim_{n}} A_n) = \lim_{k} P(B_k) \le \lim_{k} \sum_{n=k}^{\infty} P(A_n) = 0.$$

(2) 先证必要性. 由独立性知，对一切 $k, j \ge 1$ 有

$$P(\bigcap_{n=k}^{k+j} \overline{A}_n) = \prod_{n=k}^{k+j} [1 - P(A_n)],$$

任固定 k, 上式令 $j \to \infty$ 可推出

$$P(\bigcap_{n=k}^{\infty} \overline{A}_n) = \prod_{n=k}^{\infty}[1 - P(A_n)].$$

但级数 $\sum_{n=k}^{\infty} P(A_n)$ 发散蕴含上式右方无穷乘积为 0, 运用 (2.6), (2.7) 式及概率的次可加性便得证

$$P(\overline{\lim_n} A_n) = 1 - P(\underline{\lim_n} \overline{A}_n) \geq 1 - \sum_{k=1}^{\infty} P(\bigcap_{n=k}^{\infty} \overline{A}_n) = 1.$$

再证充分性. 反设 $\sum_{n=1}^{\infty} P(A_n) < +\infty$, 则由已证明的 (1) 可得 $P(\overline{\lim_n} A_n) = 0$, 矛盾. 引理得证.

此引理中 (1) 的逆命题不真. 这就是说, 如果没有独立性假设, 由 $P(\overline{\lim_n} A_n) = 0$ 不能导出 $\sum_{n=1}^{\infty} P(A_n)$ 收敛. 其证明留作练习.

由于级数 $\sum_{n=1}^{\infty} P(A_n)$ 非收敛即发散, 故由上述引理知, 独立事件列 $\{A_n\}$ 之上限集发生, 即它们中有无限多个发生的概率非 0 即 1, 从而排除了此概率取 0 至 1 之间其它数值的可能性. 事实上, 每个独立事件列 $\{A_n\}$ 的尾事件 (即由 $n \to \infty$ 时 $\{A_n\}$ 的性质所决定的事件) 的概率都是非 0 即 1 的. 概率论中将这种性质称作**0–1 律**.

如下引理包含今后将用到的几个分析事实.

引理 2.6 考虑实数列 $\{c_n\}$ 及其部分和 $s_n = \sum_{k=1}^{n} c_k$ 所组成的数列 $\{s_n\}$.

1° 若 $c_n \to c$, 则 $s_n/n \to c$;

2° 设 $b_m = \sup\{|s_{m+k} - s_m| : k \geq 1\}$, $b = \inf\{b_m : m \geq 1\}$, 则 $\sum_{n=1}^{\infty} c_n$ 收敛的充分必要条件是 $b = 0$;

3° **柯罗奈克引理** 若 $\sum_{n=1}^{\infty} c_n/n$ 收敛, 则 $s_n/n \to 0$.

证明: 1° 对任 $\varepsilon > 0$, 选 n_1 使 $n > n_1$ 时 $|c_n - c| < \varepsilon/2$, 再选 $n_2 > n_1$ 使 $n > n_2$ 时 $\sum\limits_{k=1}^{n_1} |c_k - c|/n < \varepsilon/2$. 于是当 $n > n_2$ 时有

$$\left|\frac{s_n}{n-c}\right| \le \sum_{k=1}^{n} \frac{|c_n - c|}{n}$$

$$= \sum_{k=1}^{n_1} \frac{|c_k - c|}{n} + \sum_{k=n_1+1}^{n} \frac{|c_k - c|}{n} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

2° 若 $b = 0$, 则对任意 $\varepsilon > 0$, 存在 n_1 使 $b_{n_1} < \varepsilon$. 从而对一切 $k \ge 1$ 有 $|s_{n_1+k} - s_{n_1}| < \varepsilon$. 这表明 $\sum\limits_{n=1}^{\infty} c_n$ 收敛. 逆命题证明类似.

3° 再令

$$t_0 = 0, \qquad t_n = \sum_{k=1}^{n} \frac{c_k}{k}, \qquad n \ge 1.$$

则由 $c_k = k(t_k - t_{k-1})$ 可得

$$s_{n+1} = \sum_{k=1}^{n+1} k t_k - \sum_{k=1}^{n+1} k t_{k-1} = -\sum_{k=1}^{n} t_k + (n+1) t_{n+1}, \ n \ge 1$$

于是有

$$\frac{s_{n-1}}{n+1} = -\frac{n}{n+1} \cdot \frac{1}{n} \sum_{k=1}^{n} t_k + t_{n+1}, \qquad n \ge 1. \tag{2.8}$$

现设 $\sum\limits_{n=1}^{\infty} c_n/n$ 收敛, 即其部分和序列 $\{t_n\}$ 收敛到一个有穷的极限 t, 由已证得的 1° 知序列 $\{\frac{1}{n} \sum\limits_{k=1}^{n} t_k\}$ 也收敛到 t. 由此在 (2.8) 式中令 $n \to \infty$ 便得证 $s_n/n \to 0$. 引理全都证完.

引理 2.7 对任何模不超过 1 的复数 $z_1, \cdots, z_n$ 与 $w_1, \cdots, w_n$ 有

$$|z_1 \cdots z_n - w_1 \cdots w_n| \le \sum_{k=1}^{n} |z_k - w_k|. \tag{2.9}$$

证明: 当 $n=2$ 时我们有

$$|z_1z_2 - w_1w_2| = |(z_1 - w_1)z_2 + (z_2 - w_2)w_1| \le |z_1 - w_1| + |z_2 - w_2|.$$

再用归纳法便推出 (2.9) 式对一切 $n \ge 1$ 成立. 证毕.

最后, 回忆我们在 3.4.1 节的引理 4.2 中已给出指数函数 e^{ix} 之泰勒展式的余项估计, 特别地, 对一切 $x \in \Re$ 有

$$|e^{ix} - 1| \le |x| \wedge 2; \tag{2.10}$$

$$|e^{ix} - 1 - ix| \le \frac{x^2}{2} \wedge (2|x|); \tag{2.11}$$

$$|e^{ix} - 1 - ix + \frac{x^2}{2}| \le (\frac{1}{6}|x|^3) \wedge (x^2). \tag{2.12}$$

并运用它们研究了特征函数 $f(t)$ 的泰勒级数. 如果局限于无穷小的 t, 则有如下结果.

引理 2.8 只要随机变量 ξ 相应的矩有限, 则当 $t \to 0$ 时其特征函数 $f(t)$ 有展开式

$$f(t) = 1 + itE(\xi) + o(t); \tag{2.13}$$

$$f(t) = 1 + itE(\xi) - \frac{t^2}{2}E(\xi^2) + o(t^2). \tag{2.14}$$

证明: 由 (2.11) 式知

$$|e^{itx} - 1 - itx| \le \frac{t^2x^2}{2} \wedge (2|t|\,x|),$$

对 ξ 的分布函数取积分得

$$|f(t) - 1 - itE(\xi)| \le |t|E\{(\frac{|t|}{2}\xi^2) \wedge (2|\xi|)\}.$$

当 $t \to 0$ 时, 上式右方花括号中随机变量趋于 0, 且被可积的 $2|\xi|$ 所控制, 故用勒贝格控制收敛定理得证 $E\{(\frac{|t|}{2}\xi^2)\wedge(2|\xi|)\}$ 是无穷小量. 从而 (2.13) 得证. 类似地, 以 (2.12) 代替 (2.11) 式便可证明 (2.14) 式. 证毕.

4.2.3 同分布场合的大数定律

定理 2.9 辛钦大数律 对于独立同分布随机变量列 $\{\xi_n\}$, 大数定律 (2.1) 成立的充分必要条件是 $E(\xi_n)=a$ 有限.

证明: 必要性是大数定律的定义所要求的. 只需证明充分性. 假定 $\{\xi_n\}$ 之共同的特征函数为 $f(t)$, 则由引理 2.8 知 $t\to 0$ 时有

$$f(t)=1+iat+o(t). \tag{2.15}$$

从而 $\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 的特征函数为

$$[f(\frac{t}{n})]^n=\left[1+i\,a\,\frac{t}{n}+o\,(\frac{t}{n})\right]^n.$$

对充分大的 n, 用引理 2.7 可得

$$\left|[f(\frac{t}{n})]^n-[1+i\,a\,\frac{t}{n}]^n\right|\le n\left|f(\frac{t}{n})-1-i\,a\,\frac{t}{n}\right|,$$

再用 (2.15) 式知 $n\to\infty$ 时上式右方趋于 0. 于是有

$$\lim_n[f(\frac{t}{n})]^n=\lim_n(1+i\,a\,\frac{t}{n})^n=e^{iat}.$$

据连续性定理 1.10 及定理 1.6 便得 $\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 依概率收敛到 a. 证完.

注意前述贝努里大数定律也可以由辛钦大数定律推出. 以下考虑同分布情形的强大数定律成立的充要条件. 为了了解强大数定律证明的实质, 我们还是按历史上的顺序, 先考虑贝努里场合.

定理 2.10 波雷尔强大数律 设 μ_n 为 n 重贝努里试验中成功的次数, 则 $n\to\infty$ 时有

$$\frac{\mu_n}{n}\xrightarrow{a.s.}p. \tag{2.16}$$

证明: 记 $A_n = \{|\mu_n/n - p| \geq \varepsilon\}$, 则由定理 1.1 知, (2.16) 式成立的充分必要条件是

$$P(\overline{\lim_n} A_n) = 0, \qquad 任\ \ \varepsilon < 0.$$

再根据波雷尔 - 康特里引理知，这只要证明

$$\sum_{n=1}^{\infty} P(A_n) < +\infty, \qquad 任\ \ \varepsilon < 0. \tag{2.17}$$

如果仍象证明弱大数定律那样用切贝谢夫不等式，我们只能得到

$$P(A_n) \leq \frac{1}{\varepsilon^2} D(\frac{\mu_n}{n}) = \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

这一估计太粗糙，不能满足 (2.17) 式的需要. 改用 $r = 4$ 时的马尔科夫不等式可得

$$P(A_n) \leq \frac{1}{\varepsilon^4} E|\frac{\mu_n}{n} - p|^4.$$

为计算式中的 4 阶矩，仍将 μ_n 写为 n 个独立贝努里分布随机变量之和 $\mu_n = \sum\limits_{k=1}^{n} \xi_k$. 于是有

$$E[|\frac{\mu_n}{n} - p|^4] = \frac{1}{n^4} \sum_{1 \leq i,j,k,l \leq n} E[(\xi_i - p)(\xi_j - p)(\xi_k - p)(\xi_l - p)].$$

注意 $\xi_1, \cdots, \xi_n$ 相互独立，有共同的期望 p, 故上式和号中只含形如 $E[(\xi_i - p)^4]$ 与 $E[(\xi_i - p)^2(\xi_j - p)^2]$ 的项. 最后因 $E[(\xi_i - p)^4] = pq(p^3 + q^3)$, $E[(\xi_i - p)^2(\xi_j - p)^2] = p^2q^2 (i \neq j)$, 并注意到和号中共含前者 n 项，后者 $\binom{n}{4}\binom{n}{2} = 3n(n-1)$ 项，可得到

$$E[|\frac{\mu^n}{n} - p|^4] = \frac{pq}{n^4}[n(p^3 + q^3) + 3pq(n^2 - n)] < \frac{1}{4n^2}.$$

代入马尔科夫不等式右方可得

$$P(A_n) \leq \frac{1}{4n^2\varepsilon^4},$$

这足以使 (2.17) 式成立. 至此波雷尔强大数定律得证.

下面对一般的独立同分布随机变量序列给出相应的强大数定律. 回忆定理 2.10 中对贝努里情形的证明, 本质上只用到这个独立同分布序列的四阶中心矩 $E[(\xi_k-p)^4]$ 与 $E[(\xi_k-p)^2(\xi_j-p)^2]$ 有限, 并没有用到贝努里情形的其它特性. 所以, 在四阶中心矩有限的前提下, 其证明可以照搬到一般独立同分布情形. 在如下定理中, 我们在明显减弱的条件下, 得到一般独立同分布场合的强大数定律. 证明中采用恰当地选取子序列的方法, 仅用切贝谢夫不等式就得到了满足波雷尔 - 康特里引理要求的估计式. 而且运用著名的"截尾法", 连切贝谢夫不等式中"方差有限"的要求都可以绕过. 尽管这个定理的证明所用的篇幅较大, 但是所用的技巧是值得学习的.

定理 2.11 假定 $\{\xi_n\}$ 是相互独立同分布的随机变量序列, 如果它们有有限的数学期望 $E(\xi_n)=a$, 则强大数定律成立, 即 $n\to\infty$ 时有

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{a.s.} a. \tag{2.18}$$

证明: 记部分和 $S_n=\sum\limits_{k=1}^{n}\xi_k$. 如对非负情形证明了 (2.18) 式, 那么有

$$\frac{S_n}{n}=\frac{1}{n}\sum_{k=1}^{n}\xi_k^+-\frac{1}{n}\sum_{k=1}^{n}\xi_k^- \xrightarrow{a.s.} E(\xi_1^+)-E(\xi_1^-)=a.$$

故不妨假定 $\{\xi_k\}$ 是非负随机变量列.

取每个 ξ_n 的截尾随机变量

$$\xi_n^*(\omega)=\begin{cases}\xi_n(\omega), & 若\xi_n(\omega)\le n,\\ 0, & 若\xi_n(\omega)>n.\end{cases}=\xi_n(\omega)\chi_{[0,n]}[\xi_n(\omega)]. \tag{2.19}$$

其中 $\chi_{[0,n]}(x)$ 是区间 $[0,n]$ 的示性函数. 相应的部分和记为 $S_n^*=\sum\limits_{n=1}^{n}\xi_k^*$.

对每个实数 $\theta > 1$, 取自然数列的子序列 $u_n = [\theta^n]$, 其中方括号表示其中实数的整数部分. 先考查相应的子序列 $\{S^*_{u_n}\}$. 以 $F(x)$ 表 $\{\xi_n\}$ 的共同的分布函数, 则有界随机变量 $S^*_{u_n}$ 的方差

$$D(S^*_{u_n}) = \sum_{k=1}^{u_n} D(\xi_k^*) \le \sum_{k=1}^{u_n} E[(\xi_k^*)^2]$$

$$= \sum_{k=1}^{u_n} \int_{(0,k]} x^2 dF(x) \le u_n \int_0^{+\infty} x^2 \chi_{(0,u_n]}(x) dF(x). \tag{2.20}$$

用切贝谢夫不等式便得到任 $\varepsilon > 0$ 有

$$\sum_{n=1}^{\infty} P\{\frac{1}{u_n}|S^*_{u_n} - E(S^*_{u_n})| \ge \varepsilon\} \le \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 u_n^2} D(S^*_{u_n})$$

$$\le \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \int_0^{+\infty} \frac{x^2}{u_n} \chi_{(0,u_n]}(x) dF(x). \tag{2.21}$$

对于 $x > 0$ 及 $M = 2\theta/(\theta - 1)$, 如果 N 是使 $u_n \ge x$ 的最小自然数, 则因 $y \ge 1$ 时有 $y \le 2[y]$, 所以有

$$\sum_{n=1}^{\infty} \frac{1}{u_n} \chi_{(0,u_n]}(x) = \sum_{n=N}^{\infty} \frac{1}{u_n} \le 2 \sum_{n=N}^{\infty} \theta^{-n} = M\theta^{-N} \le \frac{M}{x}.$$

将此式代入 (2.21) 式便得任 $\varepsilon > 0$ 有

$$\sum_{n=1}^{\infty} P\{\frac{1}{u_n}|S^*_{u_n} - E(S^*_{u_n})| \ge \varepsilon\} \le \frac{1}{\varepsilon^2} \int_0^{+\infty} M x dF(x) = \frac{Ma}{\varepsilon^2} < +\infty.$$

运用波雷尔 - 康特里引理可得, 任 $\varepsilon > 0$ 有

$$P\{\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} [\frac{1}{u_n}|S^*_{u_n} - E(S^*_{u_n})| \ge \varepsilon]\} = 0,$$

再由上节定理 1.1 便证得 $n \to \infty$ 时有

$$\frac{1}{u_n}[S^*_{u_n} - E(S^*_{u_n})] \xrightarrow{a.s.} 0. \tag{2.22}$$

由 $\{\xi_n^*\}$ 的定义 (2.19) 式知 $0 \le \xi_n^* \uparrow \xi$, 故据单调收敛定理可得 $E(\xi_n^*) \to E(\xi_1) = a$. 再用引理 2.6.1° 推出 $\frac{1}{n}E(S_n^*) \to a$. 联合 (2.22) 式便得到

$$\frac{1}{u_n} S_{u_n}^* \xrightarrow{a.s.} a. \tag{2.23}$$

注意到

$$\sum_{n=1}^{\infty} P\{\xi_n \ne \xi_n^*\} = \sum_{n=1}^{\infty} P\{\xi_1 > n\} \le \int_0^{+\infty} P\{\xi_1 > t\} dt = E(\xi_1) < +\infty,$$

上式中最后一个等号可参见习题 3.1.5 的第 6 或 7 题. 再次运用波雷尔 - 康特里引理便得到, 任 $\varepsilon > 0$ 有

$$P\{\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} [|\xi_n - \xi_n^*| \ge \varepsilon]\} = 0,$$

这表明 $n \to \infty$ 时 $\xi_n - \xi_n^* \xrightarrow{a.s.} 0$. 用引理 2.6.1° 知

$$\frac{1}{n}[S_n - S_n^*] \xrightarrow{a.s.} 0.$$

再联合 (2.23) 式便得

$$\frac{1}{u_n} S_{u_n} \xrightarrow{a.s.} a. \tag{2.24}$$

最后, 若 $u_n \le k \le u_{n+1}$, 则由 ξ_k 的非负性知

$$\frac{u_n}{u_{n+1}} \frac{S_{u_n}}{u_n} \le \frac{S_k}{k} \le \frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}}{u_{n+1}},$$

注意 $u_{n+1}/u_n \to \theta$. 取 $\theta = (\ell+1)/\ell$, 并令 n(及相应的 k) 趋于无穷. 由 (2.24) 式知, 以概率 1 有

$$\frac{\ell}{\ell+1} \cdot a \le \underline{\lim}_{k} \frac{S_k}{k} \le \overline{\lim}_{k} \frac{S_k}{k} \le \frac{\ell+1}{\ell} \cdot a. \tag{2.25}$$

对每个 $\ell \ge 1$, 记上式不成立的例外集为 E_ℓ, 且取 $E = \bigcup_{\ell=1}^{\infty} E_\ell$. 则 $P(E) = 0$, 而在 E 之外 (2.25) 式对一切 $\ell \ge 1$ 成立. 令 $\ell \to \infty$ 便得证 $\overline{E}$ 上恒有

$$\lim_k \frac{S_k}{k} = \lim_n \frac{1}{n} \sum_{k=1}^{n} \xi_k = a,$$

这就是要证的 (2.18) 式. 证完.

我们给出强大数定律的一个简单应用.

例 2.1 用蒙特卡洛方法计算积分 在第一章的 1.3.1 节已介绍过蒙特卡洛方法. 这就是通过设计适当的随机试验求解计算问题的方法. 大数定律提供了这种算法的严格数学基础. 考虑定积分

$$I = \int_a^b f(x)dx$$

的近似计算, 其中 $f(x)$ 为区间 $[a,b]$ 上的有界连续函数. 我们介绍两种算法, 一种基于定理 2.10, 另一种基于定理 2.11.

算法 1 不妨假定 $a=0, b=1$, 且 $0 \le f(x) \le 1$. 不然, 这一假定可通过简单的坐标变换而实现. 熟知, 积分值 I 就是单位正方形 $\Omega = [0,1]\times[0,1]$ 中, 位于曲线 $y=f(x)$ 之下曲边梯形 A 的面积. 如果随机点 (ξ,η) 在 Ω 中均匀分布, 则 I 就是事件 $\{(\xi,\eta)\in A\}$ 发生的概率. 将此试验独立重复 n 次, 以 μ_n 表上述事件发生的次数, 则由波雷尔强大数定律知, 频率 μ_n/n 几乎必然趋于概率值 I. 即以概率 1 保证, 以 μ_n/n 作为积分的近似值, 当 n 很大时可以达到任何精度要求. 不难用均匀分布随机数模拟 (ξ,η) 的 n 次独立的实验值 $(x_k, y_k), k=1,\cdots,n$. 记录 $y_k \le f(x_k)$ 发生的次数 μ_n, 便可算出 I 的近似值.

算法 2 假定 $\{\xi_n\}$ 为独立随机变量列, 每个 ξ_n 均服从区间 (a,b) 上的均匀分布. 则序列 $\{f(\xi_n)\}$ 仍是独立同分布的, 有有限的数学期望

$$E[f(\xi_n)] = \int_a^b \frac{1}{b-a} f(x)dx = \frac{I}{b-a}.$$

据定理 2.11 可知, 以概率 1 有

$$\lim_n \frac{1}{n}\sum_{k=1}^n f(\xi_k) = E[f(\xi_n)] = \frac{I}{b-a}.$$

于是可通过均匀分布随机数给出 $\{\xi_n\}$ 的样本值 $\{x_n\}$, 再用近似公式

$$I \approx \frac{b-a}{n}\sum_{k=1}^n f(x_k)$$

便可完成定积分的近似计算. 强大数定律保证, 这种算法失败的概率也是 0.

这里介绍的算法均适用于重积分的近似计算, 而且计算的精确程度与积分的维数无关. 这是蒙特卡洛方法的一大长处.

4.2.4 独立情形的强大数定律

本节将同分布条件去掉, 添加其它便于验证的充分条件, 导出一般独立随机变量序列的柯尔莫戈洛夫强大数律. 为此, 先要引进更为精细的不等式.

考虑独立随机变量序列 $\{\xi_n\}$ 及其部分和序列 $\{S_n\}$. 假定每个 ξ_n 都有有限的方差 $D(\xi_n)=\sigma_n^2$, 并设 $E(\xi_n)=0$. 对于期望不为 0 的情形, 可研究序列 $\{\xi_n-E(\xi_n)\}$ 而不失一般性. 我们已有切贝谢夫不等式

$$P\{|S_n|\geq\varepsilon\}\leq\frac{1}{\varepsilon^2}D(S_n),\qquad 任\quad \varepsilon>0.$$

下面来证明, 在 $\{\xi_n\}$ 相互独立的场合, 此不等式可以大大加强.

定理 2.12 柯尔莫戈洛夫极大不等式 设随机变量 $\xi_1,\cdots,\xi_n$ 相互独立, 它们都有零均值与有限的方差. $S_n=\sum\limits_{k=1}^{n}\xi_k$. 则对任何 $\varepsilon>0$ 有

$$P\{\max_{1\leq k\leq n}|S_k|\geq\varepsilon\}\leq\frac{1}{\varepsilon^2}D(S_n). \tag{2.26}$$

证明: 对 $k=1,\cdots,n$ 令

$$A_k=\{\omega:|S_k|\geq\varepsilon;\quad 但\ |S_j|<\varepsilon,\ j=1,\cdots,k-1\}.$$

则诸 A_k 两两不相容, 且

$$\{\max_{1\leq k\leq n}|S_k|\geq\varepsilon\}=\bigcup_{k=1}^{n}A_k.$$

于是由 $P(A_k)=E(\chi_{A_k})$ 及 A_k 的定义可推出

$$P\{\max_{1\leq k\leq n}|S_k|\geq\varepsilon\}=\sum_{k=1}^{n}P(A_k)\leq\frac{1}{\varepsilon^2}\sum_{k=1}^{n}E(S_k^2\chi_{A_k}). \tag{2.27}$$

注意到 $S_k^2\chi_{A_k}$ 是 $\xi_1,\cdots,\xi_k$ 的函数，它与 $S_n-S_k=\sum\limits_{i=k+1}^{n}\xi_i$ 相互独立．故由 3.1.2 节的定理 1.2 知

$$E[(S_n-S_k)S_k\chi_{A_k}]=E[S_n-S_k]E[S_k\chi_{A_k}]=0.$$

于是 (2.27) 式可进一步化为

$$\begin{aligned}&P\{\max_{1\le k\le n}|S_k|\ge\varepsilon\}\\ &\le\frac{1}{\varepsilon^2}\sum_{k=1}^{n}\{E(S_k^2\chi_{A_k})+2E[(S_n-S_k)S_k\chi_{A_k}]+E[(S_n-S_k)^2\chi_{A_k}]\}\\ &=\frac{1}{\varepsilon^2}E[S_n^2\sum_{k=1}^{n}\chi_{A_k}]\le\frac{1}{\varepsilon^2}E(S_n^2)=\frac{1}{\varepsilon^2}D(S_n^2).\end{aligned}$$

定理得证.

至此我们已经准备好证明独立场合的强大数定律.

定理 2.13　柯尔莫戈洛夫强大数律　设 $\{\xi_n\}$ 是独立随机变量序列，满足

$$\sum_{k=1}^{\infty}\frac{D(\xi_k)}{k^2}<+\infty,\tag{2.28}$$

则强大数定律成立，即当 $n\to\infty$ 时有

$$\frac{1}{n}\sum_{k=1}^{n}[\xi_k-E(\xi_k)]\xrightarrow{a.s.}0.\tag{2.29}$$

证明:　令 $\xi_k'=\xi_k-E(\xi_k)$, 则 $E(\xi_k')=0$, $D(\xi_k')=D(\xi_k)$. 于是条件 (2.28) 式对独立序列 $\{\xi_k'\}$ 仍成立，问题化为证明 $n\to\infty$ 时有

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k'\xrightarrow{a.s.}0.\tag{2.30}$$

按引理 2.6.2° 取 $S_n=\sum\limits_{k=1}^{n}\xi_k'/k$. 并令

$$b_m=\sup\{|S_{m+k}-S_m|\quad k=1,2,\cdots\},$$

$$b = \inf\{b_m \quad m = 1, 2, \cdots\}.$$

对任意 $\varepsilon > 0$ 及自然数 n, m, 由柯尔莫戈洛夫极大不等式 (2.26) 可得

$$P\{\max_{1\le k\le n} |S_{m+k} - S_m| \ge \varepsilon\} \le \frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n} \frac{D(\xi_k')}{k^2}. \tag{2.31}$$

注意到 $n \to \infty$ 时

$$\max_{1\le k\le n} |S_{m+k} - S_m| \uparrow b_m \ge b, \quad (m \ge 1)$$

故对一切 $m \ge 1$ 有

$$\{b > \varepsilon\} \subset \{b_m > \varepsilon\} \subset \bigcup_{n=1}^{\infty}\{\max_{1\le k\le n} |S_{m+k} - S_m| \ge \varepsilon\}.$$

于是用概率的下连续性及 (2.31) 式使得

$$P\{b > \varepsilon\} \le P\{b_m > \varepsilon\} \le \frac{1}{\varepsilon^2}\sum_{k=m+1}^{\infty} \frac{D(\xi_k')}{k^2}.$$

令 $m \to \infty$ 由条件 (2.28) 式知 $P\{b > \varepsilon\} = 0$. 取 $\varepsilon = 1/\ell$, 并记使 $b > 1/\ell$ 成立的零概集为 E_ℓ, $E = \bigcup\limits_{\ell=1}^{\infty} E_\ell$, 则 $P(E) = 0$, 而且在 E 之外, 对一切 $\ell \ge 1$ 有 $b \le 1/\ell$. 因而在 E 之外恒有 $b = 0$. 用引理 2.6 的 2° 得 $\overline{E}$ 上 $\sum\limits_{k=1}^{\infty} \xi_k'/k$ 收敛. 再用引理 2.6 的 3° 便得 $\overline{E}$ 上 $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k' = 0$, 即 (2.30) 式成立. 证完.

我们介绍了贝努里场合, 独立同分布场合及独立场合的三个强大数定律. 其中定理 2.10 只是定理 2.11 的特例, 而定理 2.11 也可由定理 2.13 推出. 这里采用由特殊到一般次序逐一地证明了它们, 一是有助于读者了解证明极限定理的要点, 二是给出独立的证明便于各类型课程选讲其中部分内容.

4.2.5 习题

1. 假定 ξ_n 有如下各种概率分布，验证 $\{\xi_n\}$ 是否满足马尔科夫条件.

(1) $P\{\xi_n = \pm 2^n\} = 1/2$;

(2) $P\{\xi_n = 0\} = 1 - 2^{-2n}$, $P\{\xi_n = \pm 2^n\} = 2^{-2n-1}$;

(3) $P\{\xi_n = \pm n^s\} = 1/2$, $(s < 1/2)$;

(4) $P\{\xi_n = 0\} = 1 - n^{-1/2}$, $P\{\xi_n = \pm n\} = (4n)^{-1/2}$;

(5) $P\{\xi_n = \pm\sqrt{\ln n}\} = 1/2$.

2. 参加集会的 n 个人将他们的帽子混放在一起，会后每人任选一顶戴上. 以 S_n 表示他们中戴对自己帽子的人的个数，求证 $\frac{1}{n}[S_n - E(S_n)] \xrightarrow{P} 0$.

3. 设随机变量列 $\{\xi_n\}$ 的方差有界，且当 $|i-j| \to \infty$ 时 ξ_i 与 ξ_j 的相关系数 $r_{ij} \to 0$. 试证对 $\{\xi_n\}$ 成立大数定律.

4. 设 $\{\xi_n\}$ 为同分布随机变量列，每个 ξ_n 有有限的方差，并且当 $|k-\ell| \geq 2$ 时 ξ_k 与 ξ_ℓ 相互独立. 试证对 $\{\xi_n\}$ 大数定律成立.

5. 设 $\{\xi_k\}$ 为独立随机变量列，它们都服从区间 $(0,1)$ 上的均匀分布，求证

$$\left(\prod_{k=1}^{n} \xi_k\right)^{\frac{1}{n}} \xrightarrow{P} c,$$

并求出 c.

6. 设 $\{\xi_n\}$ 为独立随机变量序列，每个 ξ_n 有有限的方差 σ_n^2, 试证，若 $\sigma_n^2/n \to 0$, 则对 $\{\xi_n\}$ 成立大数定律.

7. 试用引理 2.7 证明 2.3.1 节的普阿松定理.

8. 设 $\{\xi_n\}$ 为独立随机变量序列，试证 $\xi_n \xrightarrow{a.s.} 0$ 的充分必要条件是任 $\varepsilon > 0$ 有 $\sum\limits_{n=1}^{\infty} P\{|\xi_n| \geq \varepsilon\} < +\infty$.

9. 举例说明波雷尔 – 康特里引理中，命题 (1) 的逆不成立.

10. 设 $\{\xi_n\}$ 为独立同分布随机变量序列，有有限的方差 $D(\xi_n)=\sigma^2$，求证子样方差

$$S_n^2=\frac{1}{n}\sum_{k=1}^{n}(\xi_k-\overline{\xi})^2 \xrightarrow{a.s.} \sigma^2,$$

其中 $\overline{\xi}=\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 为子样均值.

11. 设 $\{\xi_n\}$ 为独立随机变量序列，满足条件 $E(\xi_n)\to 0$ 且 $\sum\limits_{n=1}^{\infty}E(\xi_n^2)/n^2<+\infty$，试证 $\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k\xrightarrow{a.s.}0$.

12. 运用强大数定律证明，对于区间 $[0,1]$ 上任何连续函数 $f(x)$, $g(x)$，只要 $0\le f(x)\le cg(x)$, $x\in[0,1]$ ($c>0$ 为常数)，就有

$$\lim_n\int_0^1\cdots\int_0^1\frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}\,dx_1\cdots dx_n=\frac{\displaystyle\int_0^1 f(x)\,dx}{\displaystyle\int_0^1 g(x)\,dx}.$$

13. 用概率方法证明如下外尔斯特拉斯定理：对区间 $[0,1]$ 上任何连续函数 $f(x)$，必存在多项式序列 $\{b_n(x)\}$，使在区间 $[0,1]$ 上一致地有 $b_n(x)\to f(x)$. [提示：取伯恩斯坦多项式 $b_n(x)=\sum\limits_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f(\frac{k}{n})$.]

4.3 中心极限定理

4.3.1 问题的提出

在 2.4.2 节介绍正态分布时曾指出，许许多多微小的偶然因素共同作用的结果必定服从正态分布. 例如，影响产品质量的因素甚多，除去产品的原材料构成以及生产工艺等主要因素外，诸如生产中能源的波动，操作者的情绪波动，生产环境中的偶然干扰，测量误差等因素都会对产品的质量指标产生影响. 于是这种质量指标总是呈现“两头小，中间大”的状态，即可认为服从正态分布. 现以随机变量

ξ_k 表示第 k 个偶然因素，那么 n 个因素的共同作用就是它们的和 $S_n = \sum_{k=1}^{n} \xi_k$．我们的问题是，当 n 趋于无穷时，这个和的概率分布是否以正态分布为极限？

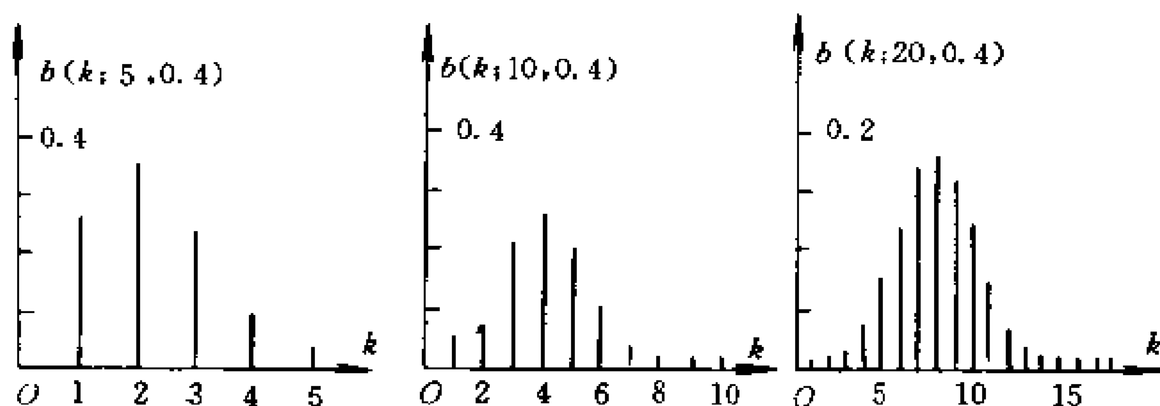


图 3.1　贝努里分布的卷积

先看两个具体例子，总设 $\{\xi_n\}$ 是相互独立的随机变量序列．如果它们有相同的贝努里分布 $b(k;1,p)$，那么其部分和 $S_n = \sum_{k=1}^{n} \xi_k$ 服从二项分布 $b(k;n,p)$．图 3.1 分别对 $n = 5, 10, 20$ 画出了 $b(k;n,0.4)$ 的密度图．易见，当 n 变大时，$b(k;n,p)$ 的图形越来越接近正态分布的密度曲线．如果每个 ξ_k 都服从区间 $(0,1)$ 上的均匀分布，我们以 $p_n(x)$ 表示和 $S_n = \sum_{k=1}^{n} \xi_k$ 的密度函数，用卷积公式可得

$$p_1(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{其它}. \end{cases} \qquad p_2(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \le x < 2, \\ 0, & \text{其它}. \end{cases}$$

$$p_3(x) = \begin{cases} x^2/2, & 0 < x < 1 \\ [x^2 - 3(x-1)^2]/2, & 1 \le x < 2, \\ [x^2 - 3(x-1)^2 + 3(x-2)^2]/2, & 2 \le x < 3, \\ 0, & \text{其它}. \end{cases}$$

图 3.2 对 $n = 1, 2, 3$ 画出了 $p_n(x)$ 的图形．由图可见，随着 n 的变大，它们也越来越接近正态密度的钟型曲线．

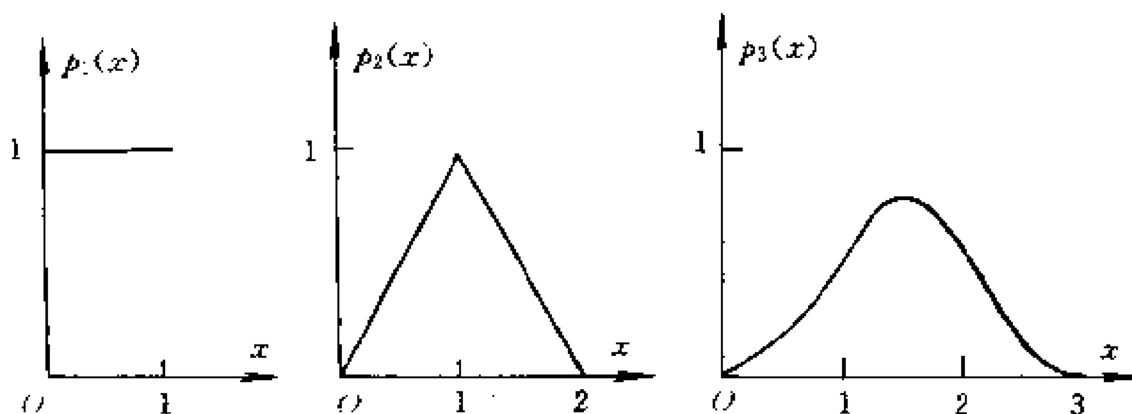


图 3.2　均匀分布的卷积

注意，对上例所述的同分布序列情形，我们有 $E(S_n)=nE(\xi_1)$ 与 $D(S_n)=nD(\xi_1)$. 故当 $E(\xi_1)$ 与 $D(\xi_1)$ 非 0 时，和 S_n 的期望与方差将随 n 趋于无穷. 就是说，S_n 的分布密度的中心将移至无穷远，且取值越来越分散. 为避免这种情形发生，我们采用在 3.2.2 节所介绍的标准化方法. 取 $\sum\limits_{k=1}^{n}\xi_k$ 的标准化随机变量

$$\zeta_n=\frac{\sum\limits_{k=1}^{n}[\xi_k-E(\xi_k)]}{\sqrt{D(\sum\limits_{k=1}^{n}\xi_k)}},\qquad n\geq 1. \tag{3.1}$$

只要每个 ξ_k 的期望与方差均有限，而且上式分母不为 0 (即非退化)，这种标准化总是可行的，且有 $E(\zeta_n)\equiv 0$ 与 $D(\zeta_n)\equiv 1$. 至此，问题化为在不同场合下讨论由 (3.1) 式定义的 ζ_n 是否以标准正态分布为其极限分布.

定义 3.1　对独立随机变量序列 $\{\xi_n\}$，如果 $E(\xi_n), D(\xi_n)$ 均有限且 $D(\xi_n)>0$ $(n\geq 1)$，而当 $n\to\infty$ 时由 (3.1) 式定义的 ζ_n 的分布函数

$$P(\zeta_n<x)\longrightarrow\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt,\qquad x\in\Re, \tag{3.2}$$

则称 $\{\xi_n\}$ 满足**中心极限定理**.

由连续性定理知，(3.2) 式等价于 ζ_n 的特征函数

$$f_n(t) = E(e^{it\zeta_n}) \longrightarrow e^{-\frac{t^2}{2}}, \qquad t \in \Re. \tag{3.3}$$

历史上，关于这个论题的第一个结果是由法国数学家德莫瓦佛于 1730 年在对称贝努里情形证明的. 在此后的大约 200 年当中，有关独立随机变量和的极限分布的讨论，一直是概率论研究的一个中心，故称作中心极限定理.

以下讨论中心极限定理成立的条件. 注意由定义 3.1 可见，每个 ξ_n 有有限的期望及有限且非 0 的方差，是 $\sum\limits_{k=1}^{n} \xi_k$ 可按 (3.1) 式标准化的前提，从而也是中心极限定理成立的必要条件.

4.3.2 同分布情形

我们将证明，添加同分布的假设之后，上节末所述的必要条件也是充分的. 就是说，独立同分布随机变量序列 $\{\xi_n\}$ 总是满足中心极限定理的.

定理 3.1 勒维－林德伯格定理 设 $\{\xi_n\}$ 是独立同分布随机变量序列，有共同的期望 $E(\xi_n) = a$ 及方差 $0 < D(\xi_n) = \sigma^2 < +\infty$，则 $\{\xi_n\}$ 满足中心极限定理.

证明: 以 $f(t)$ 表序列 $\{\xi_n - a\}$ 的共同的特征函数，则由 4.2.2 节的 (2.14) 式知，对一切 $t \in \Re$，当 $n \to \infty$ 时 $(\xi_k - a)/(\sigma\sqrt{n})$ 的特征函数可写为

$$f\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right). \tag{3.4}$$

在引理 2.7 中，取

$$z_k = f\left(\frac{t}{\sigma\sqrt{n}}\right), \qquad w_k = 1 - \frac{t^2}{2n}, \qquad k = 1, \cdots, n$$

并联合 (3.4) 式可推出，对一切 $t \in \Re$，当 $n \to \infty$ 时有

$$\left|\left[f\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n - \left(1 - \frac{t^2}{2n}\right)^n\right| \le n\left|f\left(\frac{t}{\sigma\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right)\right| \to 0.$$

至此便得证对一切 $t \in \Re$ 有

$$\lim_n \left[f\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \lim_n \left(1 - \frac{t^2}{2n}\right) = e^{-\frac{t^2}{2}},$$

即 $\{\xi_n\}$ 满足中心极限定理 (3.3) 式. 定理证完.

有了同分布场合的勒维 - 林德伯格中心极限定理, 我们可以对高尔顿钉板试验 (参见 1.1.1 节) 作出理论解释. 考虑任一指定小球的下落过程, 定义

$$\xi_k = \begin{cases} 1, & \text{若第 } k \text{ 次碰撞使小球向右,} \\ -1, & \text{若第 } k \text{ 次碰撞使小球向左.} \end{cases}$$

则 $\{\xi_k\}$ 为相互独立随机变量列, 都服从 -1 与 1 两点上的等可能分布. $E(\xi_k) = 0$, $D(\xi_k) = 1$, 满足定理 3.1 的条件. 因此, 经 n 排铁钉碰撞后小球的位置 $\eta_n = \sum\limits_{k=1}^{n} \xi_k$ 应满足

$$\lim_n P\left\{ \frac{\eta_n}{\sqrt{n}} < x \right\} = \Phi(x).$$

这便是多个小球下落后所堆积的轮廓近似为正态曲线的原因.

由定理 3.1 可导出贝努里场合的中心极限定理.

推论 3.2 德莫瓦佛 - 拉普拉斯定理 设 μ_n 为二项分布 $b(k; n, p)$ 随机变量, 则

$$\lim_n P\left\{ \frac{\mu_n - np}{\sqrt{npq}} < x \right\} = \Phi(x), \qquad x \in \Re. \tag{3.5}$$

证明: 写 $\mu_n = \sum\limits_{k=1}^{n} \xi_k$, 其中 $\{\xi_k\}$ 为独立随机变量序列, 同服从贝努里分布 $b(k; 1, p)$, $E(\xi_k) = p$, $D(\xi_k) = pq > 0$, 按 (3.1) 式所取 μ_n 的标准化 $\zeta_n = (\mu_n - np)/\sqrt{npq}$. 据定理 3.1 立得 $\{\xi_n\}$ 满足中心极限定理, 即 (3.5) 式成立. 证完.

贝努里场合的中心极限定理先由德莫瓦佛对 $p = q = 1/2$ 给出证明, 后经拉普拉斯推广到一般的 $0 < p < 1$ 情形.

4.3.3 林德伯格条件

本节讨论去掉同分布假设后中心极限定理成立的条件. 最重要的就是由林德伯格于 1922 年给出的如下充分条件.

假设 $\{\xi_n\}$ 是独立随机变量序列, 对每个 $k \geq 1$, 以 $F_k(x)$ 表 ξ_k 的分布函数, 它有有限的期望 $E(\xi_k) = a_k$ 及方差 $D(\xi_k) = \sigma_k^2$ $(\sigma_k > 0)$, 并记 $B_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$. 如果

$$\lim_n \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{(|x-a_k| \geq \tau B_n)} (x - a_k)^2 \, dF_k(x) = 0, \quad 任\ \tau > 0, \tag{3.6}$$

则称此 $\{\xi_n\}$ 满足**林德伯格条件**. 如下定理给出此条件的一个概率解释.

定理 3.3 若条件 (3.6) 成立, 则有

$$\lim_n P\left\{ \max_{1 \leq k \leq n} \left|\frac{\xi_k - a_k}{B_n}\right| \geq \tau \right\} = 0, \quad 任 \quad \tau > 0. \tag{3.7}$$

证明: 如记 $A_k = \{|\xi_k - a_k| \geq \tau B_n\}$, $1 \leq k \leq n$, 则由 (3.6) 式知

$$\begin{aligned} P\left\{ \max_{1 \leq k \leq n} |\xi_k - a_k| \geq \tau B_n \right\} &= P\left(\bigcup_{k=1}^{n} A_k \right) \\ &\leq \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n} \int_{(|x-a_k| \geq \tau B_n)} dF_k(x) \\ &\leq \frac{1}{\tau^2 B_n^2} \sum_{k=1}^{n} \int_{(|x-a_k| \geq \tau B_n)} (x - a_k)^2 \, dF_k(x) \to 0. \end{aligned}$$

定理得证.

(3.7) 式表明, 林德伯格条件 (3.6) 的一个推论是, (3.1) 式中 ζ_n 的各被加项 "依概率均匀地小". 这一点与本节开头所提到的各偶然因素都是很微小的非主导因素的要求是相吻合的. 以下是本节中最重要的中心极限定理.

定理 3.4 林德伯格定理 使条件 (3.6) 成立的独立随机变量序列 $\{\xi_n\}$ 必满足中心极限定理.

证明: 我们从林德伯格条件 (3.6) 出发, 导出 $\{\zeta_n\}$ 的特征函数列 $\{f_n(t)\}$ 满足 (3.3) 式.

对一切 $n \geq 1$ 及 $k = 1, \cdots, n$, 令 $\xi_{nk} = (\xi_k - a_k)/B_n$. 则 $E(\xi_{nk}) = 0$ 而 $D(\xi_{nk}) = \sigma_k^2/B_n^2$. 因此, 若以 F_{nk} 及 f_{nk} 分别表 ξ_{nk} 的分布函数及特征函数, 则已知的条件 (3.6) 式化为

$$\lim_n \sum_{k=1}^n \int_{(|x| \geq \tau)} x^2 \, dF_{nk}(x) = 0, \qquad \text{任 } \tau > 0. \tag{3.8}$$

而待证的 (3.3) 式变为 $n \to \infty$ 时有

$$f_n(t) = \prod_{k=1}^n f_{nk}(t) \longrightarrow e^{-\frac{t^2}{2}}, \qquad \text{任 } t \in \Re. \tag{3.9}$$

据 4.2.2 节 (2.12) 式知, 对一切 $t \in \Re$ 有

$$\left| e^{itx} - \left(1 + itx - \frac{1}{2} t^2 x^2\right) \right| \leq |tx|^2 \wedge |tx|^3,$$

注意到 ξ_{nk} 二阶矩有限, 两边对 F_{nk} 取积分得

$$\left| f_{nk}(t) - \left(1 - \frac{t^2 \sigma_k^2}{2B_n^2}\right) \right| \leq E\{|t\xi_{nk}|^2 \wedge |t\xi_{nk}|^3\}.$$

而对任何 $\tau > 0$, 上式右方不超过

$$\int_{(|x|<\tau)} |tx|^3 \, dF_{nk}(x) + \int_{(|x| \geq \tau)} (tx)^2 \, dF_{nk}(x)$$

$$\leq \tau |t|^3 \frac{\sigma_k^2}{B_n^2} + t^2 \int_{(|x| \geq \tau)} x^2 \, dF_{nk}(x).$$

由于 $\sum\limits_{k=1}^n \sigma_k^2/B_n^2 = D(\zeta_n) = 1$, 求和便推出

$$\sum_{k=1}^n \left| f_{nk}(t) - \left(1 - \frac{t^2 \sigma_k^2}{2B_n^2}\right) \right| \leq \tau |t|^3 + t^2 \sum_{k=1}^n \int_{(|x| \leq \tau)} x^2 \, dF_{nk}(x).$$

上式先令 $n\to\infty$ 再令 $\tau\downarrow 0$, 由 (3.8) 式得

$$\sum_{k=1}^{n}\left| f_{nk}(t)-(1-\frac{t^2\sigma_k^2}{2B_n^2})\right|\to 0,\qquad t\in\Re.$$

运用 4.2.2 节引理 2.7 便推出对一切 $t\in\Re$ 有

$$\left|\prod_{k=1}^{n} f_{nk}(t)-\prod_{k=1}^{n}\left(1-\frac{t^2\sigma_k^2}{2B_n^2}\right)\right|\longrightarrow 0. \tag{3.10}$$

另一方面，对任何 $\tau>0$ 有

$$\frac{\sigma_k^2}{B_n^2}=D(\xi_{nk})\le\tau^2+\int_{(|x|\ge\tau)}x^2\,dF_{nk}(x),$$

从而

$$\max_{1\le k\le n}\frac{\sigma_k^2}{B_n^2}\le\tau^2+\sum_{k=1}^{n}\int_{(|x|\ge\tau)}x^2\,dF_{nk}(x).$$

再次令 $n\to\infty$ 及 $\tau\downarrow 0$, 用 (3.8) 式得

$$\lim_{n}\left\{\max_{1\le k\le n}\frac{\sigma_k^2}{B_n^2}\right\}=0. \tag{3.11}$$

我们再次用引理 2.7 可得

$$\begin{aligned}&\left|\prod_{k=1}^{n}e^{-\frac{t^2\sigma_k^2}{2B_n^2}}-\prod_{k=1}^{n}\left(1+\frac{t^2\sigma_k^2}{2B_n^2}\right)\right|\\ &\quad\le\sum_{k=1}^{n}\left|e^{-\frac{t^2\sigma_k^2}{2B_n^2}}-1+\frac{t^2\sigma_k^2}{2B_n^2}\right|.\end{aligned} \tag{3.12}$$

注意当 $|z|\le 1/2$ 时有

$$|e^z-1-z|\le\frac{1}{2}\sum_{k=2}^{\infty}|z|^k=\frac{1}{2}|z|^2/(1-|z|)\le z^2, \tag{3.13}$$

取

$$z=\frac{t^2\sigma_k^2}{2B_n^2},$$

则由 (3.11) 式知，只要 n 充分大就有 $|z| \leq 1/2$. 故用 (3.13) 可得 (3.12) 式右方不超过

$$\frac{t^4}{4}\sum_{k=1}^{n}\left(\frac{\sigma_k^2}{B_n^2}\right)^2 \leq \frac{t^4}{4}\left(\max_{1\leq k\leq n}\frac{\sigma_k^2}{B_n^2}\right)\sum_{k=1}^{n}\frac{\sigma_k^2}{B_n^2}.$$

再由 (3.11) 知上式右方对一切 $t \in \Re$ 是趋于 0 的，联合 (3.10) 式便得对一切 t 有

$$\lim_n \prod_{k=1}^{n} f_{nk}(t) = \lim_n \prod_{k=1}^{n}\left(1-\frac{t^2\sigma_k^2}{2B_n^2}\right)$$

$$= \lim_n e^{-\frac{t^2}{2}\sum\limits_{k=1}^{n}\frac{\sigma_k^2}{B_n^2}} = e^{-\frac{t^2}{2}}.$$

这就得到了待证的 (3.9) 式. 定理证毕.

林德伯格条件的重要性将在下节作进一步的讨论. 它的缺点是不便于验证，下面给出两个便于验证的充分条件.

定理 3.5 设 $\{\xi_n\}$ 为独立随机变量列，如果存在常数列 $\{L_n\}$，使得

$$\max_{1\leq k\leq n}|\xi_k| \leq L_n, \qquad 且 \qquad \lim_n \frac{L_n}{B_n} = 0, \tag{3.14}$$

则 $\{\xi_n\}$ 满足中心极限定理.

证明: 由假设知，对于 $\tau > 0$, 存在 N, 使当 $n \geq N$ 时有 $2L_n < \tau B_n$，因此当 $n \geq N$ 时对一切 $k = 1, \cdots, n$ 有 $\Omega = \{|\xi_k - a_k| < \tau B_n\}$，故 $n \geq N$ 时有

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{(|x-a_k|<\tau B_n)}(x-a_k)^2\,dF_k(x)$$

$$= \frac{1}{B_n^2}\sum_{k=1}^{n}\int_{\Re}(x-a_k)^2\,dF_k(x) = 1.$$

这表明林德伯格条件 (3.6) 式成立，用定理 3.4 便得 $\{\xi_n\}$ 满足中心极限定理. 证完.

定理 3.6 李雅普诺夫定理 设 $\{\xi_n\}$ 为独立随机变量列，如果存在正数 δ，使

$$\lim_n \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|\xi_k - a_k|^{2+\delta} = 0, \tag{3.15}$$

则 $\{\xi_n\}$ 满足中心极限定理.

证明: 只要验证林德伯格条件成立. 实际上对任何 $\tau > 0$ 有

$$\begin{aligned}
&\lim_n \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{(|x-a_k| \geq \tau B_n)} (x - a_k)^2 \, dF_k(x) \\
\leq &\lim_n \frac{1}{B_n^2 (\tau B_n)^\delta} \sum_{k=1}^{n} \int_{(|x-a_k| \geq \tau B_n)} |x - a_k|^{2+\delta} \, dF_k(x) \\
\leq &\lim_n \frac{1}{\tau^\delta} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} \int_{\Re} |x - a_k|^{2+\delta} \, dF_k(x) = 0.
\end{aligned}$$

定理证完.

条件 (3.15) 通常称作**李雅普诺夫条件**. 我们给出应用上述二定理的例.

例 3.1 设 $\{\xi_n\}$ 为独立随机变量序列，对每个 $k \geq 1, \xi_k$ 服从区间 $(-\sqrt{k}, \sqrt{k})$ 上的均匀分布，试证 $\{\xi_n\}$ 满足中心极限定理.

解 1: 易见对 $k \geq 1$ 有 $E(\xi_k) = 0, D(\xi_k) = k/3$, 从而 $B_n^2 = n(n+1)/6$. 取 $L_n = \sqrt{n}$, 则条件 (3.14) 成立. 故 $\{\xi_n\}$ 满足中心极限定理.

解 2: 取 $\delta = 2$, 则 $B_n^{2+\delta} = [n(n+1)/6]^2$，而

$$\sum_{k=1}^{n} E|\xi_k - a_k|^{2+\delta} = \sum_{k=1}^{n} \int_{-\sqrt{k}}^{\sqrt{k}} \frac{x^4}{2\sqrt{k}} \, dx = \frac{1}{30} n(n+1)(2n+1).$$

因此当 $n \to \infty$ 时有

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|\xi_k - a_k|^{2+\delta} = \frac{6(2n+1)}{5n(n+1)} \longrightarrow 0.$$

即李雅普诺夫条件 (3.15) 成立，所以 $\{\xi_n\}$ 满足中心极限定理.

4.3.4 费勒条件

在已经给出的中心极限定理成立的充分条件中，林德伯格条件是最弱的了. 既使在同分布场合的定理 3.1 中，所设的条件也比林德伯格条件强. 事实上，此时假定共同的分布满足 $0 < \sigma^2 < +\infty$，这个条件蕴含

$$\begin{aligned}&\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{(|x-a_k|\geq \tau B_n)}(x-a_k)^2\,dF_k(x)\\&=\frac{1}{\sigma^2}\int_{(|x-a|\geq \tau\sigma\sqrt{n})}(x-a)^2\,dF(x)\to 0,\end{aligned}$$

即林德伯格条件成立. 但是从中心极限定理的定义 3.1 可知，$0 < D(\xi_n) + \infty$ 是中心极限定理的必要条件，所以林德伯格条件在同分布场合已经是中心极限定理的充分必要条件了. 但是去掉同分布假设后，林德伯格条件已不必要，且看如下的例.

例 3.2 假定 $\{\xi_k\}$ 为独立随机变量序列，而每个 ξ_k 服从 $N(a_k, \sigma_k^2)$ 分布，$0 < \sigma_k^2 < +\infty$. 那么，每个 $\zeta_n = \sum\limits_{k=1}^{n}(\xi_k - a_k)/B_n^2$ 均服从 $N(0,1)$ 分布，从而 $\{\xi_n\}$ 满足中心极限定理.

以下取 $\sigma_k^2 = 2^{-k}\ (k \geq 1)$，则 $B_n^2 = 1 - 2^{-n} < 1\ (n \geq 1)$. 故对任何 $\tau > 0$ 有

$$\begin{aligned}&\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{(|x-a_k|\geq \tau B_n)}(x-a_k)^2\,dF_k(x)\\&\geq \int_{(|x-a_1|>\tau)}(x-a_1)^2\,dF_1(x).\end{aligned}$$

由正态分布性质知，右方是一个正数，且与 n 无关，所以林德伯格条件不成立.

回顾定理 3.4 的证明，我们在林德伯格条件 (3.6) 之下，导出了一个中间结果 (3.11) 式. 如果把方差 σ_k^2 当作被加项 ξ_k 对和式的“贡

献”的话，那么 (3.11) 式表明，与总和相比，各被加项的贡献是“可忽略地小”. 而我们所举的使林德伯格条件不成立的例 3.2, 恰恰是破坏了 (3.11) 式. 事实上此时有

$$\max_{1\le k\le n}\frac{\sigma_k^2}{B_n^2}=\frac{2^{-1}}{1-2^{-n}}\longrightarrow\frac{1}{2}.$$

于是人们猜想，如果将林德伯格条件中由 (3.11) 式所代表的那部分性质“分离”出去，它是否会成为中心极限定理的充分必要条件呢？费勒严格地证明了这一点， (3.11) 式也被称作**费勒条件**.

定理 3.7 费勒定理 设 $\{\xi_n\}$ 为独立随机变量列， $E(\xi_k)=a_k$ 与 $D(\xi_k)=\sigma_k^2$ 均有限，且 $B_n^2=\sum\limits_{k=1}^{n}\sigma_n^2>0\ (n\ge1)$. 那么中心极限定理 (3.3) 式及费勒条件 (3.11) 式同时成立的充分必要条件是林德伯格条件 (3.6) 式成立. 即有

$$(3.3)+(3.11)\iff(3.6).$$

证明： 条件的充分性已由定理 3.4 给出. 下证必要性，即假设 $\{\xi_n\}$ 满足中心极限定理 (3.3) 及费勒条件 (3.11), 往证林德伯格条件 (3.13) 成立. 仍记 $\xi_{nk}=(\xi_k-a_k)/B_n$, 并以 $F_{nk}(x)$ 及 $f_{nk}(t)$ 分别表示它的分布函数与特征函数. 在以下的讨论中，总是任意固定 $t\in\Re$.

由 4.2.2 节 (2.10) 式知

$$|f_{nk}(t)-1|\le 2E\{1\wedge|t\xi_{nk}|\},$$

故对任何 $\varepsilon>0$ 有

$$\begin{aligned}|f_{nk}(t)-1|&\le 2\int_{(|x|\ge\varepsilon)}dF_{nk}(t)+2\int_{(|x|<\varepsilon)}|tx|\,dF_{nk}(x)\\&\le 2P\{|\xi_{nk}|\ge\varepsilon\}+2\varepsilon|t|.\end{aligned}$$

再由切贝谢夫不等式得

$$P\{|\xi_{nk}|\ge\varepsilon\}\le\frac{1}{\varepsilon^2}\frac{\sigma_k^2}{B_n^2}.\tag{3.16}$$

于是对 $\varepsilon > 0$ 有

$$\max_{1\le k\le n}|f_{nk}(t)-1| \le \frac{2}{\varepsilon^2}\left\{\max_{1\le k\le n}\frac{\sigma_k^2}{B_n^2}\right\}+2\varepsilon|t|.$$

上式先令 $n\to\infty$，再令 $\varepsilon\downarrow 0$，用 (3.11) 式便得到

$$\max_{1\le k\le n}|f_{nk}(t)-1| \longrightarrow 0. \tag{3.17}$$

再据上节 (2.11) 式可推出

$$|f_{nk}(t)-1| \le \frac{t^2\sigma_k^2}{B_n^2}, \quad 及 \quad \sum_{k=1}^{n}|f_{nk}(t)-1|\le t^2,$$

联合 (3.17) 式便得

$$\sum_{k=1}^{n}|f_{nk}(t)-1|^2 \le \left\{\max_{1\le k\le n}|f_{nk}(t)-1|\right\}\sum_{k=1}^{n}|f_{nk}(t)-1|$$

$$\le t^2\left\{\max_{1\le k\le n}|f_{nk}(t)-1|\right\} \longrightarrow 0. \tag{3.18}$$

因为当 $|z|\le 1$ 时 $|e^{z-1}| = e^{Rez-1}\le 1$，故可对 $z_k = \exp\{f_{nk}(t)$ $-1\}$ 与 $w_k = f_{nk}(t)$ 运用 4.2.2 节引理 2.7 推出

$$\left|\exp\left\{\sum_{k=1}^{n}[f_{nk}(t)-1]\right\}-\prod_{k=1}^{n}f_{nk}(t)\right|$$

$$\le \sum_{k=1}^{n}|\exp\{f_{nk}(t)-1\}-f_{nk}(t)|. \tag{3.19}$$

注意 (3.16) 及 (3.11) 式表明 $\{\xi_{nk}\}$ 依概率 (从而依分布) 收敛到 0, 故 k 充分大时有 $|f_{nk}(t)-1|\le 1/2$. 对 $z = f_{nk}(t)-1$ 用不等式 (3.13) 便知 (3.19) 式的右端不超过 (3.18) 式的左端，从而有

$$\lim_n \exp\left\{\sum_{k=1}^{n}[f_{nk}(t)-1]\right\} = \lim_n \prod_{k=1}^{n}f_{nk}(t).$$

但是中心极限定理 (3.3) 成立蕴含上式右端的极限为 $e^{-t^2/2}$, 至此我们得到

$$\lim_n \exp\left\{\sum_{k=1}^n [f_{nk}(t)-1]\right\} = e^{-\frac{t^2}{2}}.$$

考虑相应的模，由 $|e^z| = e^{Rez}$ 并取对数便得

$$\lim_n \sum_{k=1}^n Re[f_{nk}(t)-1] = -\frac{t^2}{2},$$

上式可改写为

$$\sum_{k=1}^n E\left\{\cos(t\xi_{nk}) - 1 + \frac{1}{2}t^2\xi_{nk}^2\right\} \longrightarrow 0. \tag{3.20}$$

这里 Rez 表示复数 z 的实部. 由 $\cos x \geq 1 - x^2/2$ 知，上式花括号中随机变量非负，从而 (3.20) 式左端不小于

$$\sum_{k=1}^n \int_{(|x|\geq\tau)} \left(\frac{t^2x^2}{2} - 2\right) dF_{nk}(x)$$

$$\geq \left(\frac{t^2}{2} - \frac{2}{\tau^2}\right) \sum_{k=1}^n \int_{(|x|\geq\tau)} x^2\, dF_{nk}(x). \tag{3.21}$$

注意以上讨论均对一切 $t \in \Re$ 成立. 因为对任何 $\tau > 0$, 总存在 t 使 $\frac{t^2}{2} - \frac{2}{\tau^2} > 0$, 联合 (3.20) 及 (3.21) 式便得证林德伯格条件 (3.8) 式成立. 证毕.

4.3.5 应用举例

例 3.3 二项分布的正态近似 根据德莫瓦佛－拉普拉斯中心极限定理，可用正态分布近似计算参数 n 较大时的二项分布随机变量 μ_n 的有关概率，即有近似公式

$$P\left\{\frac{\mu_n - np}{\sqrt{npq}} < x\right\} \approx \Phi(x). \tag{3.22}$$

与 2.3.1 节中二项分布的普阿松近似相比，这里不需要“p 很小”这一要求．我们来算一个具体的例子．

假定某车间有 200 台同型机床，每台开动需供电 1 千瓦．如果每台机床的开动率均为 $p = 0.6$，问供电多少千瓦可使供电不足的概率在千分之一以下．

解：显然，这个车间中开动的机床数 μ_n 是 $b(k;200,0.6)$ 分布的随机变量，问题是求 r，使

$$P\{\mu_n \le r\} \ge 0.999.$$

采用近似公式 (3.22)，问题化为求 r，使

$$P\left\{\frac{\mu_n - np}{\sqrt{npq}} \le \frac{r-np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{r-120}{\sqrt{48}}\right) \ge 0.999.$$

查附表 III 可得，只需取

$$\frac{r-120}{\sqrt{48}} = 3.1 \qquad \text{即} \qquad r = 141$$

便可满足要求．就是说，只向这个车间供电 141 千瓦，则由于供电不足而影响生产的可能性小于千分之一，相当于在 8 小时的工作日中可能有半分钟受影响，这在一般工厂中是允许的．

例 3.4 正态随机数的产生 在随机模拟中，经常用到正态随机数．前面的 2.7.6 节已介绍过产生正态随机数的方法．有了同分布场合的勒维－林德伯格中心极限定理，可以直接用均匀分布随机数生成正态分布随机数．设 $\{\xi_n\}$ 是独立随机变量列，它们都服从区间 $(0,1)$ 上的均匀分布．由定理 3.1 知，$\sqrt{\frac{12}{n}}\sum\limits_{k=1}^{n}(\xi_k - \frac{1}{2})$ 渐近服从标准正态分布．通常取 $n = 12$，将均匀分布随机数 $\xi_1,\cdots,\xi_{12}$ 代入 $\sum\limits_{k=1}^{12}\xi_k - 6$，经检验知其正态性已能满足需要．

例 3.5 误差分析 人们对客观存在的量进行测量或计算时，所得的数值与真值之间必然会有误差．比如测量或计算精确到小数点后

5 位，第 6 位四舍五入，则一次测量或计算的误差 (界) 为 0.5×10^{-5}. 如果以 ξ_k 表示第 k 次计算的误差，则进行 n 次运算的累积误差 (界) 可用传统的绝对值不等式来估计:

$$\left|\sum_{k=1}^{n}\xi_k\right|\leq\sum_{k=1}^{n}|\xi_k|\leq n\cdot0.5\times10^{-5}. \tag{3.23}$$

这显然是一种保守的估计. 在实际应用中，由于多次运算中有舍入相抵，故实际误差应当小得多. 事实上，每个误差应当是区间 $(-0.5\times10^{-5},0.5\times10^{-5})$ 上均匀分布的随机变量，而累积误差是独立和 $\sum\limits_{k=1}^{n}\xi_k$. 运用 "$3\sigma$ 原则", 以 99.97% 的概率保证

$$\left|\sum_{k=1}^{n}\xi_k\right|\leq3\sqrt{n\cdot10^{-10}/12}=\frac{\sqrt{3n}}{2}\cdot10^{-5}. \tag{3.24}$$

以 $n=10000$ 为例，这个估计仅是 (3.23) 式中估计的 1/60.

历史上，德国著名数学家高斯首先将正态分布成功地用于误差分析，以致人们把正态分布称作**高斯分布**. 而法国的德莫瓦佛与拉普拉斯远在这之前就引用正态分布研究了贝努里场合的中心极限定理，他们却没能得到高斯这样的荣誉. 可见数学理论运用到生产实践中之后会产生巨大的影响.

4.3.6 习题

1. 反复抛一均匀硬币，为保证正面出现的频率在 0.4 至 0.6 之间的概率不小于 90%, 试用 (1) 切贝谢夫不等式； (2) 二项分布的正态近似确定至少要抛多少次？

2. 假定生男孩的概率近似等于 0.515，求 10000 个婴孩中男孩不多于女孩的概率.

3. 现有一大批种子，其中良种占 1/6. 今从中任取出 6000 粒，问能以 99% 的概率保证其中良种的比例与 1/6 相差多少？

4. 某保险公司售出人寿保险 10000 份，每份收保险金 12 元，并规定在保期内死亡者可得到 1000 元的赔付金. 如果每个持保单者在

保险期内死亡的概率 $p=0.006$, 求保险公司 (1) 亏本; (2) 盈利达 40000 元; (3) 盈利达 60000 元; (4) 盈利达 80000 元 的概率.

5. 设 $\{\xi_n\}$ 为独立随机变量列, 对 $n\geq 1, \xi_n$ 有如下分布 (1) $P\{\xi_n=\pm\sqrt{n}\}=1/2$; (2) $P\{\xi_n=0\}=1/3, P\{\xi_n=\pm n^{\alpha}\}=1/3, (\alpha>0)$. 验证李雅普诺夫条件是否成立?

6. 用中心极限定理证明: $n\to\infty$ 时 $e^{-n}\sum\limits_{k=0}^{n} n^k/k!\to 1/2$.

7. 设 $\eta=0.\xi_1\xi_2\cdots$ 是区间 $[0,1]$ 上均匀分布的随机变量, 试证序列 $\{\xi_n\}$ 满足中心极限定理.

8. 设独立随机变量序列 $\{\xi_n\}$ 满足中心极限定理, 试证 $\{\xi_n\}$ 满足大数定律的充分条件是 $\frac{1}{n^2}\sum\limits_{k=1}^{n} D(\xi_k)\to 0$.

9. 设 $\{\xi_n\}$ 为独立随机变量列, 每个 ξ_n 有分布 $P\{\xi_n=\pm n^{\alpha}\}=1/2$, 试证 (1) 当 $\alpha>-1/2$ 时 $\{\xi_n\}$ 满足中心极限定理; (2) 已知 $\alpha>0$ 时, $\{\xi_n\}$ 服从大数定律的充分必要条件是 $\alpha<1/2$.

10. 设有 n 个口袋, 对 $k=1,\cdots,n$, 第 k 个口袋中有 1 个白球与 $k-1$ 个黑球,

(1) 现从每袋中各任取 1 球, 以 ξ_n 表取得的白球总数, 证明 $\{\xi_n\}$ 渐近正态分布.

(2) 现从每袋中有放回地任取 2 球, 若 2 球全白则认为是成功的. 以 η_n 表成功总次数, 求证 $\{\eta_n\}$ 不是渐近正态分布的.

11. 设随机变量列 $\{\xi_n\}$ 独立同分布, $E(\xi_n)=0, D(\xi_n)=1$. 试证

$$(1)\ \eta_n=\sqrt{n}\frac{\xi_1+\cdots+\xi_n}{\xi_1^2+\cdots+\xi_n^2};\qquad (2)\ \zeta_n=\frac{\xi_1+\cdots+\xi_n}{\sqrt{\xi_1^2+\cdots+\xi_n^2}}$$

均为渐近正态分布的.

12. 设每个 ξ_n 有分布

$$P\{\xi_n=\pm n\}=\frac{1}{2n^2},\qquad P\{\xi_n=0\}=1-\frac{1}{n^2}.$$

求证独立随机变量序列 $\{\xi_n\}$ 不满足中心极限定理.

13. 试证费勒条件 (3.11) 式等价于

$$B_n^2 \to +\infty \qquad 且 \qquad \sigma_n^2/B_n^2 \to 0.$$

14. 设 $\{\xi_n\}$ 为独立随机变量序列，每个 ξ_n 有 $N(0,\sigma_n^2)$ 分布，其中 $\sigma_1^2 = 1$, 对 $n > 1$, $\sigma_n^2 = nB_{n-1}^2$. 试证 $\{\xi_n\}$ 满足中心极限定理但不满足费勒条件.

部 分 习 题 答 案

习题 1.2.4

2. (1) $\binom{n+r-1}{r}$, (2) $\binom{r-1}{n-1}$ **3.** $\frac{r}{r+b}$ **4.** $\frac{\binom{k-1}{i-1}r^i b^{k-i}}{(r+b)^k}$

5. $\frac{\binom{n}{k}\binom{n-k}{2m-2k}2^{2m-2k}}{\binom{2n}{2m}}$ **6.** $\frac{1}{2}$ **7.** (1) $\frac{A_{10}^8}{10^8}$, (2) 10^{-7}, (3) 10^{-8},

(4) $\frac{\binom{8}{3}9^5}{10^8}$, (5) 0.9^8 **8.** (1) $\frac{n!}{n^n}$, (2) $\left(\frac{n-1}{n}\right)^n$ **9.** $\frac{1}{\binom{n+r-1}{r}}$, $\frac{\binom{n}{r}}{\binom{n+r-1}{r}}$,

$\frac{\binom{r-m+n-2}{r-m}}{\binom{n+r-1}{r}}$ **10.** (1) $\frac{9}{16}$ (2) $\frac{19}{64}$ **11.** (1) $\frac{\binom{4}{1}}{\binom{52}{5}}$, (2) $\frac{\binom{4}{1}\binom{8}{1}}{\binom{52}{5}}$,

(3) $\frac{\binom{13}{1}\binom{48}{1}}{\binom{52}{5}}$, (4) $\frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}$, (5) $\frac{\binom{4}{1}\binom{13}{5}}{\binom{52}{5}}$, (6) $\frac{\binom{9}{1}(4^5-4)}{\binom{52}{5}}$,

(7) $\frac{\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2}{\binom{52}{5}}$, (8) $\frac{\binom{13}{2}\binom{4}{2}^2\binom{44}{1}}{\binom{52}{5}}$, (9) $\frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3}{\binom{52}{5}}$,

(10) $\frac{\left[\binom{13}{5}-9\right](4^5-4)}{\binom{52}{5}}$ **12.** $\frac{9!\,48!}{54!}$ **13.** $\frac{b!(b+1)!}{(b+r)!(b-r+1)!}$

14. (1) $\frac{\binom{N}{n}}{N^n}$, (2) $\frac{\binom{N+n-1}{n}}{N^n}$ **15.** $\frac{1}{2^{n-1}}$

习题 1.3.3

1. $\frac{1}{3}$ **2.** $\frac{3}{4}$ **3.** $<\frac{10}{9}$ **4.** 0.27 **5.** (1) 0.68, (2) 0.64, (3) 0.46

6. (1) $\frac{13}{24}$, (2) $\frac{1}{48}$ **7.** (1) $\frac{1}{4}$, (2) $\frac{2}{3}$

习题 1.4.5

1. $\frac{19}{20}$ **2.** $1-\frac{(N-1)^{k-1}}{N^k}$ **3.** $1-\frac{2}{(2n-1)!}$ **4** $\frac{5^n-4^n}{6^n}$

5. (1) $\frac{2^k-1}{3^{k-1}}$, (2) $\frac{2^{k-1}-2}{3^{k-1}}$ 6. $\frac{4^n+6\cdot 2^n-4(3^n+1)}{4^n}$

7. $1-\frac{\binom{8}{n}+\binom{5}{n}-\binom{4}{n}}{\binom{10}{n}}$ 8. $\frac{l_1+l_2+l_3}{\pi a}$ 9. $\sum_{i=0}^{N}(-1)^i\binom{n}{i}(1-\frac{i}{N})^n$

10. $\sum_{i=1}^{n}(-1)^{i-1}\frac{n!}{i!}$ 15. (2) $\frac{2}{3}$, (3) $1-2^{-n}$

习题 1.5.4

1. $\frac{6}{7}$ 2. $\frac{1}{2}$ 3. $\frac{2}{3}$ 4. $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$, 0 6. $\frac{1}{30}$ 7. $\frac{\binom{4}{3}\binom{6}{k-4}}{\binom{10}{k-1}}\cdot\frac{1}{(11-k)}$,

8. $\frac{b}{r+b+s}$ 9. $\frac{1-\alpha_1}{3-\alpha_1}$, $\frac{1}{3-\alpha_1}$, $\frac{1}{3-\alpha_1}$ 10. 0.75, 0.25 11. 2^{-n}

12. $\frac{1}{2}+(c-\frac{1}{2})(2p-1)^{n-1}$ 13. $q_n=\frac{2}{3}+\frac{1}{3}(-\frac{1}{2})^n$, $p_n=r_n=\frac{1}{4}q_{n-1}$

14. (1) $\left(\frac{r-2}{r-1}\right)^{n-1}$, (2) $\frac{(r-2)(r-3)\cdots(r-n)}{(r-1)^{n-1}}$, (3) $\frac{1}{r}(1-$

$[\frac{1}{1-r}]^{n-2})$ 15. $\frac{1}{2\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{4}\right)^{n-1}-\left(\frac{1-\sqrt{5}}{4}\right)^{n-1}\right]$

习题 1.6.5

3. $\frac{1}{2}$, $\frac{1}{2}$ 7. (1)0.36, (2) 0.91 8. 0.61 9. 0.33 10. (1) $\prod_{i=1}^{n}(1-p_i)$,

(2) $1-\prod_{i=1}^{n}(1-p_i)$, (3) $\prod_{i=1}^{n}(1-p_i)+\sum_{i=1}^{n}[p_i\prod_{j\neq i}(1-p_j)]$

习题 2.1.5

1. $F(\ln x)$, $x>0$ 2. $F(x)=0$, $x\le a$; $=P(A)$, $a<x\le b$;

$=1$, $x>b$ 4. (1)$pq^{k-1}+qp^{k-1}$, $k=2,3,\cdots$ (2) $\frac{p}{1+p}+\frac{q}{1+q}$

5. (1) $F(x)=\frac{x^2}{R^2},\ 0<x<R$, (2) $\frac{5}{9}$ 6. (1) $F(x)=0,\ x\le 0;$ $=\frac{x^2}{2},\ \ 0<x\le 1;\ \ =-\frac{x^2}{2}+2x-1,\ 1<x\le 2;\ \ =1,\ x>2,$ (2) 0.66 7. (1) e^{-1}, (2) e^{-1}

习题 2.2.5

1. (1) 0.988, (2) 0.206, (3) 4 2. $\left(\frac{25}{36}\right)^{n-1}\frac{11}{36}$ 3. 请 5 名顾问

4. $\frac{1}{2}[1+(q-p)^n]$ 5. $\binom{i-1}{3}0.4^{i-4}0.6^4,\ i=4,\cdots,7$, 三局两胜制 6. $\binom{2n-r-1}{n-1}2^{r-2n+1}$ 8. $\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r}$ 9. $\frac{p_i}{p_i+p_j}$

习题 2.3.4

1. $1-2.5e^{-1}\approx 0.08$ 2. 102只 3. 16台 4. 0.957 5. 0.933

6. $e^{-\frac{1}{2}}\approx 0.61$ 7. $1-0.04(1+2\ln 5)\approx 0.831$

习题 2.4.5

1. $\frac{3}{5}$ 4. (1) 2.14, (2) 3.92 5. 0.1587, 0.3413, 0.3413, 0.1587

6. (1) 穿行, (2) 绕行 7. $N(0,1)$ 8. $g(k;1-e^{-\lambda})$ 10. (1) e^{-1}, (2) $e^{-\frac{1}{2}}$

习题 2.5.5

1. 联合分布 $p_{ij}=\frac{1}{4i}$; 当 $i=1,2,3,4$ 且 $j\le i$; $=0$, 其它

2. 对$n\ge 1$, (1) $p_{nn}=0.36\times 0.24^{n-1}$, $p_{n(n-1)}=0.4\times 0.24^{n-1}$, (2) $p_{n\cdot}=0.76\times 0.24^{n-1}$, $p_{\cdot n}=0.456\times 0.24^{n-1}$, $p_{\cdot 0}=0.4$

3 (1)

$$F(x,y)=\begin{cases}1, & x,y\ge\frac{\pi}{2},\\ \frac{1}{2}(1+\sin x-\cos x), & 0<x<\frac{\pi}{2},y\ge\frac{\pi}{2},\\ \frac{1}{2}(1+\sin y-\cos y), & x\ge\frac{\pi}{2},0<y<\frac{\pi}{2},\\ \frac{1}{2}\{1+\sin x+\sin y-\sin(x+y)\}, & 0<x,y<\frac{\pi}{2},\\ 0, & \text{其它}.\end{cases}$$

(2) $p_2(y)=\dfrac{1}{2}(\cos y+\sin y)$ **6.** (1) $\dfrac{3}{2}$, (2) $\dfrac{23}{128}$ **7.** (1) 联合分布为 $p_{mn}=p^2q^{n-2}$, $1\le m<n$, (2) $p_{m\cdot}=q^{m-1}p$, $m\ge 1$, $p_{\cdot n}=(n-1)p^2q^{n-2}$, $n\ge 2$ **8.** (1) $p_1(x)=6(x-x^2)$, $x\in(0,1)$, (2) $\dfrac{7}{4}-\sqrt{2}$ **9.** $1-e^{-\frac{\lambda^2}{2(1-r^2)}}$ **11.** (1) $p_1(x)=e^{-x}$, $x>0$, $p_2(y)=\dfrac{1}{(1+y)^2}$, $y>0$, (2) $p_{13}(x,z)=ze^{-x-z}$, $x,z>0$

习题 2.6.3

1. $j=1$: $\begin{pmatrix}1 & 2 & 3 & 4\\ \frac{12}{25} & \frac{6}{25} & \frac{4}{25} & \frac{3}{25}\end{pmatrix}$, $j=2$: $\begin{pmatrix}2 & 3 & 4\\ \frac{6}{13} & \frac{4}{13} & \frac{3}{13}\end{pmatrix}$, $j=3$: $\begin{pmatrix}3 & 4\\ \frac{4}{7} & \frac{3}{7}\end{pmatrix}$, $j=4$: $\begin{pmatrix}4\\ 1\end{pmatrix}$ **2.** 点$1,2,\cdots,n-1$上等可能分布 **3.** $\dfrac{\sin(x+y)}{\cos y+\sin y}$, $0<x<\dfrac{\pi}{2}$ **4.** (1) $24y-48y^2$, $0<y<\dfrac{1}{2}$, (2) $4-8x$, $0<x<\dfrac{1}{2}$ **5.** $\lambda e^{-\lambda x}$, $x>0$ **7.** (1) 独立, (2) 不独立

习题 2.7.7

1. (1) 服从巴斯卡分布 $f(k;2,p)$, (2) $pq^{k-1}(2-q^{k-1}-q^k)$

2. (1) $\begin{pmatrix}0 & 2 & 3 & 4 & 5\\ \frac{2}{16} & \frac{7}{16} & \frac{1}{16} & \frac{3}{16} & \frac{3}{16}\end{pmatrix}$, (2) $\begin{pmatrix}1 & 2 & 3\\ \frac{4}{16} & \frac{6}{16} & \frac{6}{16}\end{pmatrix}$, (3) $\begin{pmatrix}0 & 1 & 4\\ \frac{3}{16} & \frac{7}{16} & \frac{6}{16}\end{pmatrix}$ **3.** (1) $p(k;\lambda_1-\lambda_2)$, (2) $b(k;n,\dfrac{\lambda_1}{\lambda_1+\lambda_2})$

4. (1) $\frac{1}{y\sqrt{2\pi}}e^{-\frac{\ln^2 y}{2}}$, $y>0$, (2) $\frac{e^{-\frac{1}{2y}}}{\sqrt{2\pi y^3}}$, $y>0$ **5.** $\lambda=\sum_{i=1}^{n}\lambda_i$的指数分布 **6.** (1) $F(y)=\begin{cases}\int_{-\infty}^{0}p(x)dx+\int_{\frac{1}{y}}^{+\infty}p(x)dx, & y>0,\\ \int_{-\infty}^{0}p(x)dx, & y=0,\\ \int_{\frac{1}{y}}^{0}p(x)dx, & y<0.\end{cases}$

(2) $F(y)=\sum_{k=-\infty}^{+\infty}\int_{k\pi-\frac{\pi}{2}}^{k\pi+\mathrm{tg}^{-1}y}p(x)dx$, (3) $\int_{-y}^{y}p(x)dx$, $y>0$

7. $p_{\xi+\eta}(x)=\begin{cases}x, & 0<x<1,\\ 2-x, & 1\le x<2.\end{cases}$ **9.** $p(x)=\begin{cases}\frac{1}{2}, & 0<x<1,\\ \frac{1}{2x^2}, & x\ge 1.\end{cases}$

10. $p(r)=\frac{2a-2r}{a^2}$, $0<r<a$ **11.** $p(z)=z^{-\frac{1}{2}}-1$, $0<z<1$

12. $p(u,v)=\frac{1}{2u^2v}$, $u\ge 1$且$\frac{1}{u}\le v\le u$, $p_1(u)=\frac{\ln u}{u^2}$, $u\ge 1$,

$p_2(v)=\begin{cases}\frac{1}{2}, & 0<v<1,\\ \frac{1}{2v^2}, & v\ge 1.\end{cases}$ **13.** 独立同$N(0,2)$分布 **17.** ρ, θ独立,

$p_\rho(r)=4r(1-r^2)$, $0<r<1$, $p_\theta(t)=\frac{1}{2\pi}$, $0<t<2\pi$.

18. $3u^2(1+u)^{-4}$, $u>0$ **19.** $p(y)=\begin{cases}\frac{2}{3}, & 0<y<1,\\ \frac{2}{3y^3}, & y\ge 1\end{cases}$ **20.**

(1) $F(x)=\begin{cases}\frac{x}{2}, & 0<x\le 2,\\ 1, & x>2.\end{cases}$ (2) $F(x)=\begin{cases}\frac{x}{2}, & 0<x\le\frac{1}{2},\\ x-\frac{1}{4}, & \frac{1}{2}<x\le 1,\\ \frac{x}{2}+\frac{1}{4}, & 1<x\le\frac{3}{2},\\ 1, & x\ge\frac{3}{2}.\end{cases}$

(3) $F(x)=\begin{cases}\frac{1}{2}+\frac{x}{2}, & 0<x<1\\ 1, & x\ge 1.\end{cases}$

习题 3.1.6

1. (1) $\frac{n+1}{2}$ (2) n **2.** μ **3.** $\frac{r}{\lambda}$ **4.** $\frac{(n-1)a}{n+1}$ **8.** 1 **9.** $\frac{\ln 2}{\pi}+\frac{1}{2}$

11. $\frac{(1-p)(3-p)}{p(2-p)}$　12 $\frac{nr}{r+b}$　13. 1　14. $\frac{mn}{2}$　15. $\frac{r+b+1}{r+1}$

习题 3.2.7

1. (1) $\frac{n^2-1}{12}$, (2) $n(n-1)$　2. $2\lambda^2$　3. $\frac{5}{8}$　4. 0　5. npq　6. (1) $\frac{m(n^2-1)}{12}$　(2) $\frac{m(n-m)(n+1)}{12}$　7. 1　8. $\begin{pmatrix} \frac{11}{144} & \frac{-1}{144} & 0 \\ \frac{-1}{144} & \frac{11}{144} & 0 \\ 0 & 0 & 2 \end{pmatrix}$

9. $1-\frac{m}{n}$　12. $\frac{mn}{N}$　13. η, η^2　14. (1) $\frac{3}{5}e^{-2}+\frac{2}{5}e^{-3}$, $\frac{4}{5}e^{-2}+\frac{9}{5}e^{-3}$　(2) $\frac{1}{2}$, $\frac{1}{16}$　15. $\frac{\eta^4}{4}$　16. $\frac{n}{2}$　19. $\frac{\lambda^k}{k!}$

20. $n(n+2)\cdots(n+2k-2)$

习题 3.3.4

1. (1) $\frac{(e^{s\lambda}-1)e^{-\lambda}}{1-e^{-\lambda}}$, (2) $\frac{1-(sq)^{N+1}}{1-q^{N+1}}$　2. $\left(\frac{ps}{1-qs}\right)^r$, $\frac{r}{p}$, $\frac{rq}{p^2}$

3. $s^n G(s^m)$　4. $\frac{11}{200}$　5. $e^{-\lambda t(1-\sum_{n=1}^{\infty} p_n s^n)}$, $\lambda t\sum_{n=1}^{\infty} np_n$　6. (1) $\frac{G(s)}{1-s}$, (2) $\frac{1}{2}\{G(\sqrt{s})+G(-\sqrt{s})\}$　7. $\binom{2n}{n+k}4^{-n}$　8. $\frac{1-qs}{(1-qs+ps)(1-s)}$

习题 3.4.5

1. $\left(\frac{pe^{it}}{1-qe^{it}}\right)^r$, $\frac{r}{p}$　2. $\left(1-i\frac{t}{\lambda}\right)^{-r}$, $\lambda^{-k}r(r+1)\cdots(r+k-1)$

3. (1), (2), (5), (6), (7) 是特征函数　7. (1) $\frac{2}{t^2}(1-\cos t)$, (2) $\frac{1}{\pi x^2}(1-\cos x)$　8. $\frac{1}{1+t^2}$　11. $C(0,1)$　16. (1) $\frac{\sin s}{s}\frac{\sin t}{t}+$

$+\frac{c}{s^4t^2}\{(3s^2-6)\sin s-(s^3-6s)\cos s\}(\sin t-t\cos t)$

$$-\frac{c}{s^2t^4}\{(3t^2-6)\sin t-(t^3-6t)\cos t\}(\sin s-s\cos s),$$

(2) $f_1(s)=\dfrac{\sin s}{s}$, $f_2(t)=\dfrac{\sin t}{t}$, (3) $f_{\xi+\eta}(t)=\left(\dfrac{\sin t}{t}\right)^2$

习题 3.5.4

1. (1) $\dfrac{1}{6\sqrt{3}\pi}\exp\{-\dfrac{1}{54}(9x^2-6xy+4y^2+6x-8y+4)\}$,

(2) $\exp(it-2s^2-3st-\dfrac{9}{2}t^2)$ **3.** $\dfrac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}}$, $x>0$ **4.** $\begin{pmatrix}1 & \frac{2}{\pi}\\ \frac{2}{\pi} & 1\end{pmatrix}$

5. $b_{11}=b_{22}$ **6.** $\dfrac{ac+bd}{\sqrt{(a^2+c^2)(b^2+d^2)}}$ **7.** $N(\bar{0},B)$分布，其中

$B=\begin{pmatrix}m & m\\ m & n\end{pmatrix}$ **8.** (1) $N(0,\dfrac{\sigma^2}{n})$, (2) $\dfrac{1}{\sqrt{n}}$

习题 4.2.5

1. (1), (4) 不满足, (2), (3), (5) 满足. **5.** $c=-1$

习题 4.3.6

1. (1) 250, (2) 73 **2.** 0.00135 **3.** 0.0125 **4.** (1) 0;

(2) 0.994739; (3) 0.5; (4) 0.005261 **5.** (1) 与(2) 均成立.

中 英 人 名 对 照

巴拿赫	Banach	惠更斯	Huygens
贝叶斯	Bayes	雅可比	Jacobi
贝努里	Bernoulli	辛钦	Khintchine
伯恩斯坦	Bernstein	柯尔莫戈洛夫	Kolmogorov
贝特朗	Bertrand	柯罗奈克	Kronecker
波雷尔	Borel	拉普拉斯	Laplace
蒲丰	Buffon	勒贝格	Lebesgue
柯西	Cauchy	勒维	Lévy
康特里	Cantelli	林德伯格	Lindeberg
切贝谢夫	Chebyshev	李雅普诺夫	Lyapunov
德莫瓦佛	De Moivre	马尔柯夫	Markov
笛卡儿	Descartes	巴斯卡	Pascal
杜勃	Doob	普阿松	Poisson
欧拉	Euler	瑞利	Rayleigh
法都	Fatou	卢瑟福	Rutherford
费勒	Feller	施瓦兹	Schwarz
费尔马	Fermat	斯蒂尔吉斯	Stieltjes
富里埃	Fourier	斯特灵	Stirling
高尔顿	Galton	泰勒	Taylor
高斯	Gauss	威布尔	Weibull
海莱	Helly	外尔斯特拉斯	Weierstrass

参考书目

[1] 王梓坤， 概率论基础及其应用，科学出版社， 1976.

[2] 复旦大学编， 概率论（第一册），人民教育出版社， 1979.

[3] 中山大学编， 概率论及数理统计（上册），人民教育出版社， 1980.

[4] 华东师范大学编， 概率论与数理统计习题集， 人民教育出版社，1982.

[5] 严士健等， 概率论基础，科学出版社， 1982.

[6] 周概容， 概率论与数理统计，高等教育出版社， 1984.

[7] 汪嘉冈， 现代概率论基础，复旦大学出版社， 1988.

[8] 严加安， 测度与积分，陕西师范大学出版社， 1988.

[9] 何声武等译， 概率论习题集，上海翻译出版公司， 1989.

[10] Ross,S., A First Course in Probability, Macmillan Publishing Co., New York, (中译本，李漳南，杨振明译，概率论初级教程，人民教育出版社， 1981.)

[11] Chung,K.L., A Course in Probability Theory, Academic Press, New York, (中译本，刘文，吴让泉译，概率论教程，上海科学技术出版社， 1989.)

[12] Loève,M., Probability Theory, 4th edition, Springer-Verlag, New York, 1977.

[13] Chow,Y.S. & Teicher,H., Probability Theory, Springer- Verlag, New York, 1978.

[14] Billingsley, E., Probability and Measure, John Wiley & Sons, New York, 1986.

附表 I 常用分布表

分布名称	参数	概率密度	期望	方差	母函数	特征函数	备注
退化分布（一点分布）	c	$\begin{pmatrix} c \\ 1 \end{pmatrix}$	c	0		e^{ict}	关于参数 c 有再生性
贝努里分布（两点分布）	p $0<p<1$	$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$	p	pq	$q+ps$	$q+pe^{it}$	是 $n=1$ 时的二项分布
二项分布 $b(k;n,p)$	$n\geq 1$ $0<p<1$	$\binom{n}{k} p^k q^{n-k}$, $k=0,1,\cdots,n$	np	npq	$(q+ps)^n$	$(q+pe^{it})^n$	关于 n 有再生性
几何分布 $g(k;p)$	p $0<p<1$	$q^{k-1}p$, $k=1,2,\cdots$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{ps}{1-qs}$	$\frac{pe^{it}}{1-qe^{it}}$	是 $r=1$ 的巴斯卡分布；无记忆
巴斯卡分布（负二项分布）$f(k;r,p)$	r,p r 为自然数 $0<p<1$	$\binom{k-1}{r-1} p^r q^{k-r}$ $k=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left(\frac{ps}{1-qs}\right)^r$	$\left(\frac{pe^{it}}{1-qe^{it}}\right)^r$	关于 r 有再生性
普阿松分布 $p(k;\lambda)$	λ $\lambda>0$	$\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\cdots$	λ	λ	$e^{\lambda(s-1)}$	$e^{\lambda(e^{it}-1)}$	关于 λ 有再生性

续上表

分布名称	参　数	概率密度	期望	方　差	母函数	特征函数	备　注
超几何分布	M,N,n 均为自然数	$\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ $k=0,1,\cdots,n$	$\frac{nM}{N}$	$\frac{nM}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$			
均匀分布 $U(a,b)$	a,b $a<b$	$\frac{1}{b-a}$; $a<x<b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$		$\frac{e^{itb}-e^{ita}}{it(b-a)}$	
正态分布 $N(a,\sigma^2)$	a,σ^2 $\sigma>0$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$, $-\infty<x<+\infty$	a	σ^2		$e^{iat-\frac{1}{2}\sigma^2t^2}$	关于 a,σ^2 有再生性
指数分布	λ $\lambda>0$	$\lambda e^{-\lambda x}$, $x>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$		$\left(1-\frac{it}{\lambda}\right)^{-1}$	是 $\Gamma(\lambda,1)$ 分布；无记忆
伽马分布 $\Gamma(\lambda,r)$	λ,r $\lambda,r>0$	$\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$, $x>0$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$		$\left(1-\frac{it}{\lambda}\right)^{-r}$	关于参数 r 再生性
χ^2 分布 $\chi^2(n)$	n $n\geq 1$	$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}$, $x>0$	n	$2n$		$(1-2it)^{-\frac{n}{2}}$	关于 n 有再生性；是 $\Gamma(\frac{1}{2},\frac{n}{2})$ 分布；是正态平方和的分布

续上表

分布名称	参数	概率密度	期望	方差	母函数	特征函数	备注
柯西分布 $C(\lambda,\mu)$	λ,μ $\lambda>0$	$\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$, $-\infty<x<+\infty$	不存在	不存在		$e^{i\mu t-\lambda\|t\|}$	是 $t(1)$ 分布；是正态之商的分布；关于 λ,μ 有再生性
瑞利分布	σ $\sigma>0$	$\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$, $x>0$	$\sqrt{\frac{\pi}{2}}\sigma$	$\left(2-\frac{\pi}{2}\right)\sigma^2$			是 $\chi^2(2)$ 随机变量之算术根的分布
对数正态分布	a,σ $\sigma>0$	$\frac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{(\ln x-a)^2}{2\sigma^2}}$, $x>0$	$e^{a+\frac{\sigma^2}{2}}$	$e^{2a+\sigma^2}(e^{\sigma^2}-1)$			是 e^{ξ} 的分布，其中 ξ 为 $N(a,\sigma^2)$ 随机变量
威布尔分布	α,λ $\alpha,\lambda>0$	$\alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}}$, $x>0$	$\Gamma(\frac{1}{\alpha}+1)\cdot\lambda^{-\frac{1}{\alpha}}$	$\lambda^{-\frac{2}{\alpha}}\left[\Gamma(\frac{2}{\alpha}+1)-\left(\Gamma(\frac{1}{\alpha}+1)\right)^2\right]$			
拉普拉斯分布	λ,μ $\lambda>0$	$\frac{1}{2\lambda}e^{-\frac{\|x-\mu\|}{\lambda}}$, $-\infty<x<+\infty$	μ	$2\lambda^2$		$\frac{e^{i\mu t}}{1+\lambda^2t^2}$	
t 分布 $t(n)$	n $n\geq 1$	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$, $-\infty<x<+\infty$	0 $(n>1)$	$\frac{n}{n-2}$ $(n>2)$			是独立的正态与 χ 分布之商的分布
F 分布 $F(m,n)$	m,n $m,n\geq 1$	$\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{B(\frac{m}{2},\frac{n}{2})}x^{\frac{m}{2}-1}$ $(n+mx)^{-\frac{m+n}{2}}$, $x>0$	$\frac{n}{n-2}$ $(n>2)$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $(n>4)$			是独立 χ^2 分布之商的分布
贝塔分布	α,β $\alpha,\beta>0$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$, $0<x<1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$			

附表 II　普阿松分布数值表

$$p(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}$$

k	λ							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0	.904837	.818731	.740818	.670320	.606531	.548812	.496585	.449329
1	.090484	.163746	.222245	.268128	.303265	.329287	.347610	.359463
2	.004524	.016375	.033337	.053626	.075816	.098786	.121663	.143785
3	.000151	.001092	.003334	.007150	.012636	.019757	.028388	.038343
4	.000004	.000055	.000250	.000715	.001580	.002964	.004968	.007669
5	–	.000002	.000015	.000057	.000158	.000356	.000696	.001227
6			.000001	.000004	.000013	.000036	.000081	.000164
7	–	–	–		.000001	.000003	.000008	.000019
8			–	–		–	.000001	.000002

续上表

k	λ							
	0.9	1.0	1.5	2.0	2.5	3.0	3.5	4.0
0	.406570	.367879	.223130	.135335	.082085	.049787	.030197	.018316
1	.365913	.367879	.334695	.270671	.205212	.149361	.150091	.073263
2	.164661	.183940	.251021	.270671	.256516	.224042	.184959	.146525
3	.049398	.061313	.125510	.180447	.213763	.224042	.215785	.195367
4	.011115	.015328	.047067	.090224	.133602	.168031	.188812	.195367
5	.002001	.003066	.014120	.036089	.066801	.100819	.132169	.156293
6	.000300	.000511	.003530	.012030	.027834	.050409	.077098	.104196
7	.000039	.000073	.000756	.003437	.009941	.021604	.038549	.059540
8	.000004	.000009	.000142	.000859	.003106	.008102	.016865	.029770
9	–	.000001	.000024	.000191	.000863	.002701	.006559	.013231
10		–	.000004	.000038	.000216	.000810	.002296	.005292
11	–		–	.000007	.000049	.000221	.000730	.001925
12		–	–	.000001	.000010	.000055	.000213	.000642
13		–	–	–	.000002	.000013	.000057	.000197
14	–	–	–	–	–	.000003	.000014	.000056
15	–	–		–	–	.000001	.000003	.000015
16	–		–	–		–	.000001	.000004
17		–		–	–		–	.000001

续上表

k	λ							
	4.5	5.0	5.5	6.0	7.0	8.0	9.0	10.0
0	.011109	.006738	.004087	.002479	.000912	.000335	.000123	.000045
1	.049990	.033690	.022477	.014873	.006383	.002684	.001111	.000454
2	.112479	.084224	.061813	.044618	.022341	.010735	.004998	.002270
3	.168718	.140374	.113323	.089235	.052129	.028626	.014994	.007567
4	.189808	.175467	.155819	.133853	.091226	.057252	.033737	.018917
5	.170827	.175467	.171401	.160623	.127717	.091604	.060727	.037833
6	.128120	.146223	.157117	.160623	.149003	.122138	.091090	.063055
7	.082363	.014445	.123449	.137677	.149003	.139587	.117116	.090079
8	.046329	.065278	.084871	.103258	.130337	.139587	.131756	.112599
9	.023165	.036266	.051866	.068838	.101405	.124077	.131756	.125110
10	.010424	.018133	.028526	.041303	.070983	.099262	.118580	.125110
11	.004264	.008242	.014263	.022529	.045171	.072190	.097020	.113736
12	.001599	.003434	.006537	.011262	.026350	.048127	.072765	.094780
13	.000554	.001321	.002766	.005199	.014188	.029616	.050376	.072908
14	.000178	.000472	.001087	.002288	.007094	.016924	.032384	.052077
15	.000053	.000157	.000398	.000891	.003311	.009026	.019431	.034718
16	.000015	.000049	.000137	.000336	.001448	.004513	.010960	.021699
17	.000004	.000014	.000044	.000118	.000596	.002124	.005768	.012764
18	.000001	.000004	.000014	.000039	.000232	.000944	.002893	.007091
19	—	.000001	.000004	.000012	.000085	.000397	.001370	.003732
20	—	—	.000001	.000004	.000030	.000159	.000617	.001866
21	—	—	—	.000001	.000010	.000061	.000264	.000889
22					.000003	.000022	.000108	.000404
23	—	—	—	—	.000001	.000008	.000042	.000176
24	—	—	—	—	—	.000003	.000016	.000073
25		—	—	—	—	.000001	.000006	.000029
26	—	—	—				.000002	.000011
27	—	—	—	—	—	—	.000001	.000004
28	—					—	—	.000001
29	—	—		—			—	.000001

附表 III 标准正态分布数值表

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
2	.5793	.5832	.5871	.5910	.5943	.5987	.6026	.6064	.6103	.6141
3	.6179	.6217	.6255	.6293	.6631	.6368	.6106	.6443	.6480	.6517
4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1	.8543	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
3	.9032	.9049	.9066	.9085	.9099	.9115	.9131	.9147	.9162	.9177
4	.9192	.9507	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
1	.9821	.9826	.9831	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
6	.9953	.9955	.9956	9957	.9959	.9960	.9961	.9962	.9963	.9964
7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

$\Phi(4.0)=0.999968329$　　$\Phi(5.0)=0.9999997133$　　$\Phi(6.0)=0.999999999$

附表 IV 随 机 数 表

01839	96423	24878	82651	66566	14778	76797	14780	13300	87071
68086	26432	46901	20849	89768	81536	86645	12659	92259	57102
39064	66432	84673	40027	32832	61362	98947	96067	64760	64584
25669	26422	41407	44048	37937	63904	15766	66134	75470	66520
61117	94305	26766	25940	39972	22209	71500	64568	91402	42116
87917	77341	42206	35126	74087	99547	81817	42607	43808	76655
62797	56170	86324	88072	76222	36086	84637	93161	76038	65855
95876	55293	18988	27354	26575	08625	40801	59920	29841	80150
29888	88604	67917	18708	18912	82271	65424	69774	33641	54262
73577	12908	30883	18317	28290	35797	05998	41688	34952	37888
27958	30134	04024	86385	29880	99730	55536	84855	29080	09250
90999	49127	20014	59931	06115	20542	18059	02008	73708	83517
18845	49618	02304	51038	20655	58727	28168	15475	56942	53389
94824	78171	84610	82834	09922	25417	44137	48413	25555	21246
35605	81263	39667	47358	56873	56307	61607	49518	89656	20103
33362	64270	01638	92477	66969	98420	04880	45585	46565	04102
88720	82765	34476	17032	87589	40836	32427	70002	70663	88863
39475	46473	23219	53416	94970	25832	69975	94884	19661	72828
06990	67245	68350	82948	11398	42878	80287	88267	47363	46634
40980	07391	58745	25774	22937	80059	39911	96189	41151	14222
83974	29092	65831	38857	50490	83765	55657	14361	31720	57375
33339	31926	14883	24413	59744	92351	97473	89286	35931	04110
31662	25388	61642	34072	81249	35648	56891	69352	48373	45578
93526	70765	10592	04542	76463	54328	02349	17247	28865	14777
20492	38391	91132	21999	59516	81652	27195	48223	46751	22923
04153	53381	79401	21433	83035	92350	36693	31238	59649	91754
05520	91962	04739	13092	97662	24822	94730	06496	35090	04822
47498	87637	99016	71060	88824	71013	18735	20286	23153	72924
23167	49323	45021	33132	12544	41035	80780	45393	44812	12515
23792	14422	15059	45799	22716	19792	09983	74353	68668	30429
85900	98275	32388	52390	16815	69298	82732	38480	73817	32523
42559	78985	05300	22164	24369	54224	35083	19687	11052	91491
14349	82674	66523	44133	00697	35552	35970	19124	63318	29686

后　　记

我在南开大学数学系 (现数学科学学院) 为各专业学生讲授概率论课程多年. 由于每年都给几个班开课, 听课的学生早已超过千人. 特别是从 86 年以来, 我校招收了数学专业试点班. 领导对教学工作很重视, 学生的素质较高, 我们也为提高教学质量作了许多努力.

这些年, 我先后使用王梓坤教授, 李贤平教授的书作教材 (见参考书目 [1] 与 [2]). 同时也参考了国内外的不少概率论教科书. 只是近年来这些优秀教材已经买不到, 给教学带来一定困难. 现在我有机会编写本书, 希望它在今后几年内能为学生提供一些方便.

与典型的概率论教材一样, 本书包含事件与概率, 随机变量, 数字特征与特征函数, 极限定理共四章. 重点在介绍概率论的直观背景和解决初等概率问题的基本方法, 为学习专业课程打下基础. 本教材不假定学生已学过实分析或其它测度论课程. 但是众所周知, 没有这些基本知识, 要阐明概率的数学理论是不大可能的. 所以本书也试图简明通俗地介绍这方面的一些必备知识, 除有关测度扩张的定理之外, 基本做到 "自圆其说". 书中这些内容均已用星号标出 (但是它们并不难). 有些读者可以先记住其中的一些结论, 而略去其证明, 在以后的应用中逐步加深理解.

多年来, 我在概率论教学中比较注重深入浅出, 论证条理清楚. 在科学地安排例题与习题上也作出一些努力. 希望这些特点在本书中有所体现.

由于本人水平所限, 错误之处可能很多, 望不吝指正.

杨振明　　1998 年 3 月于南开大学

后　记

我在南开大学数学系 (现数学科学学院) 为各专业学生讲授概率论课程多年. 由于每年都给几个班开课, 听课的学生早已超过千人. 特别是从 86 年以来, 我校招收了数学专业试点班. 领导对教学工作很重视, 学生的素质较高, 我们也为提高教学质量作了许多努力.

这些年, 我先后使用王梓坤教授, 李贤平教授的书作教材 (见参考书日 [1] 与 [2]). 同时也参考了国内外的不少概率论教科书. 只是近年来这些优秀教材已经买不到, 给教学带来一定困难. 现在我有机会编写本书, 希望它在今后几年内能为学生提供一些方便.

与典型的概率论教材一样, 本书包含事件与概率, 随机变量, 数字特征与特征函数, 极限定理共四章. 重点在介绍概率论的直观背景和解决初等概率问题的基本方法, 为学习专业课程打下基础. 本教材不假定学生已学过实分析或其它测度论课程. 但是众所周知, 没有这些基本知识, 要阐明概率的数学理论是不大可能的. 所以本书也试图简明通俗地介绍这方面的一些必备知识, 除有关测度扩张的定理之外, 基本做到 “自圆其说”. 书中这些内容均已用星号标出 (但是它们并不难). 有些读者可以先记住其中的一些结论, 而略去其证明, 在以后的应用中逐步加深理解.

多年来, 我在概率论教学中比较注重深入浅出, 论证条理清楚. 在科学地安排例题与习题上也作出一些努力. 希望这些特点在本书中有所体现.

由于本人水平所限, 错误之处可能很多, 望不吝指正.

杨振明　　1998 年 3 月于南开大学